

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 15, 2005, 10:06:35 ; Search time 147 Seconds  
(without alignments)  
4241.070 Million cell updates/sec

Title: AF334735  
Perfect score: 1641  
Sequence: 1 TCGCCCTTCTCGCGCGCG.....AAAAAAAAAAAAAAAAAAAA 954

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 2753750

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DRV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/AF334735/runat\_15022005\_094945\_10184/app\_query.fasta.1.1095  
-DB=Published Applications AA -QFWI=fastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -LOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=50 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USPR=AF334735 @CGN 1.1 199 @runat\_15022005\_094945\_10184  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCT05\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	121.5	7.4	697	15	US-10-181-638A-2	Sequence 2, Appli
2	118.5	7.2	382	16	US-10-408-765A-1431	Sequence 1431, Ap
3	114.5	7.0	103	10	US-09-764-891-3819	Sequence 3819, Ap
4	110.5	6.6	298	14	US-10-017-161-2194	Sequence 2194, Ap
5	110.5	6.6	298	15	US-10-292-798-1840	Sequence 1840, Ap
6	108.5	6.6	511	16	US-10-437-963-148275	Sequence 148275,
7	108	6.6	403	14	US-10-369-947A-38	Sequence 38, Appl
8	107.5	6.6	805	15	US-10-369-493-152	Sequence 152, Appl
9	104	6.3	580	15	US-10-424-599-237561	Sequence 237561,
10	103.5	6.3	532	15	US-10-114-270-178	Sequence 178, App
11	103.5	6.3	535	15	US-10-425-114-64123	Sequence 64123, A
12	103	6.3	197	13	US-10-101-487-51	Sequence 51, Appl
13	103	6.3	197	13	US-10-101-487-114	Sequence 114, App
14	103	6.3	238	14	US-10-163-668-264	Sequence 264, App
15	103	6.3	252	9	US-09-778-927A-42	Sequence 42, Appl
16	102.5	6.2	582	16	US-10-437-963-145629	Sequence 145629,
17	102	6.2	267	9	US-09-778-927A-43	Sequence 43, Appl
18	102	6.2	376	14	US-10-156-761-9889	Sequence 9889, Ap
19	101	6.2	390	15	US-10-108-260A-4549	Sequence 4549, Ap
20	101	6.2	647	9	US-09-815-242-11218	Sequence 11218, A
21	101	6.2	647	15	US-10-282-122A-58476	Sequence 58476, A
22	101	6.2	649	15	US-10-467-248-3	Sequence 3, Appli
23	101	6.2	3067	10	US-09-949-029-18	Sequence 18, Appl
24	100.5	6.1	1487	16	US-10-437-963-161300	Sequence 161300,
25	100.5	6.1	2468	16	US-10-755-889-615	Sequence 615, App
26	99.5	6.1	743	14	US-10-087-454-53	Sequence 53, Appl
27	99	6.0	498	15	US-10-282-122A-72013	Sequence 72013, A
28	99	6.0	521	15	US-10-425-114-70443	Sequence 70443, A
29	99	6.0	670	15	US-10-363-616-339	Sequence 339, App
30	99	6.0	775	10	US-09-934-455-450	Sequence 450, App
31	99	6.0	1537	16	US-10-437-963-183719	Sequence 183719,
32	98.5	6.0	426	13	US-10-087-192-1197	Sequence 1197, Ap
33	98	6.0	273	15	US-10-424-599-166145	Sequence 166145,
34	98	6.0	478	15	US-10-434-599-208282	Sequence 208282,
35	98	6.0	480	9	US-09-347-331-9	Sequence 9, Appli
36	98	6.0	916	16	US-10-408-765A-1222	Sequence 1222, Ap
37	97.5	5.9	197	15	US-10-424-599-174175	Sequence 174175,
38	97.5	5.9	574	16	US-10-437-963-202437	Sequence 202437,
39	97.5	5.9	620	16	US-10-437-963-158544	Sequence 158544,
40	97.5	5.9	749	15	US-10-092-900A-20	Sequence 20, Appl
41	97	5.8	484	15	US-10-369-493-6247	Sequence 6247, Ap
42	97	5.9	995	15	US-10-094-749-2626	Sequence 2626, Ap
43	97	5.9	1261	13	US-10-147-268-2	Sequence 2, Appli
44	97	5.9	1261	14	US-10-338-279-2	Sequence 2, Appli
45	96.5	5.7	438	14	US-10-017-161-2198	Sequence 2198, Ap

#### ALIGNMENTS

RESULT 1  
US-10-181-638A-2  
Sequence 2, Application US/10181638A  
Publication No. US20030207393A1  
GENERAL INFORMATION:  
APPLICANT: Naaby-Hansen, Soren  
APPLICANT: Wolkowicz, Michael  
APPLICANT: Mandal, Arabinda  
APPLICANT: Buer, Sen  
APPLICANT: Herr, John C  
TITLE OF INVENTION: CBP86, A Sperm Specific Protein  
FILE REFERENCE: 00492-07  
CURRENT APPLICATION NUMBER: US/10181,638A  
PRIOR FILING DATE: 2002-10-15  
CURRENT FILING DATE: 2000-01-19  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 2  
LENGTH: 697  
TYPE: PRT  
ORGANISM: Homo sapiens

## US-10-181-638A-2

Alignment Scores:  
Pred. No.: 0.00339 Length: 697  
Score: 121.50 Matches: 35  
Percent Similarity: 46.03% Conservative: 23  
Best Local Similarity: 27.78% Mismatches: 43  
Query Match: 7.40% Indels: 25  
DB: 15 Gaps: 3

## AF334735 (1-954) x US-10-181-638A-2 (1-697)

QY 169 ATTCCACAGGATTTGGGAATCTTCTGAAGGGCTGACACGGCGAGATCTCGAGAGCAA 228  
Db 10 ValProTyrGlyLeuLysThrLeuLeuGluGlyIleSerArgAlaValLeuLysThrAsn 29  
QY 229 CCGGACATATACCAAGCTTTTCAGAGCTTATTTTGAGAGCTTCTAGAGAAAGAGAG 288  
Db 30 ProSerAsnIleAsnGlnPheAlaAlaIlePheGlnGluLeuThrMetTyrArgGly 49  
QY 289 AAAACCACTTTGATCCAGCAGATGGGGGAGTAAAGTAGAGCGGCTTCTATACCAAT 348  
Db 50 AsnThrThrMetAsp-----IleLysAspLeuValLysGlnPhe 62  
QY 349 CATGCATTCGAGGAGCAAGAACCCATCGAGAAAGTGATCTTAAACAAAGAA----- 399  
Db 63 HisGlnIleLysValGluLysTrpSerGluGlyThrThrProGlnLysLeuGluCys 82  
QY 400 -----GAGTCTCAGATA-----TCT 414  
Db 83 LeuLysGluProGlyLysThrSerValGluSerLysValProThrGlnMetGluLysSer 102  
QY 415 GGGAGGAGGAGAGACATCAGTCACCATCTAGACTCTTCGAGGAGATAGGAAAAA 474  
Db 103 ThrAspThrAspGluAspAsnValThrArgThrGluTyrSerAspLysThrThrGlnPhe 122  
QY 475 GAAGAGGTGCTGCTGTC 492  
Db 123 ProSerValTyrAlaVal 128

## RESULT 2

US-10-408-765A-1431  
; Sequence 1431, Application US/10408765A  
; Publication No. US20040101874A1

## GENERAL INFORMATION:

; APPLICANT: Ghosh, Soumitra S.  
; APPLICANT: Fahy, Eoin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibson, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Glenn, Gary M.  
; APPLICANT: Warnock, Dale E.  
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
; FILE REFERENCE: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME  
; CURRENT APPLICATION NUMBER: US/10/408,765A  
; CURRENT FILING DATE: 2003-04-04  
; NUMBER OF SEQ ID NOS: 3077  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1431  
; LENGTH: 382  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-408-765A-1431

Alignment Scores:  
Pred. No.: 0.00916 Length: 382  
Score: 118.50 Matches: 38  
Percent Similarity: 41.26% Conservative: 21  
Best Local Similarity: 26.57% Mismatches: 53  
Query Match: 7.22% Indels: 31  
DB: 16 Gaps: 4

## AF334735 (1-954) x US-10-408-765A-1431 (1-382)

QY 157 ACCCACTACCGAATTTCCACAGGATTTGGGAATCTTCTGAAGGGCTGACACGGCGAGATT 216  
Db 2 SerHisIleGlnIleProProGlyLeuThrGlnLeuLeuGlnGlyTyrThrValGluVal 21  
QY 217 CTGAGAGAGCAACGGGACATATACCAAGCTTTTCAGAGCTTATTTTGAGAGCTTCTTA 276  
Db 22 LeuArgGlnGlnProProAspLeuValGluPheAlaValGluTyrPheThrArgLeuArg 41  
QY 277 GAGAAAAGAGAGAAAACCAACTTTGATCCAGCAGATGGGGGAGTAAAGTAGAGAGACGC 336  
Db 42 GluAlaArgAlaProAlaSerValLeuProAla-----AlaThrProArg 56  
QY 337 TTCTATAACAATCATGATTCGAGAGCAAGAACCCATCGAGAAAGTGATCTTAAACAA 396  
Db 57 GlnSerLeuGlyHisProProGlnProGlyProAspArgValAlaAspAlaLysGly 76  
QY 397 GAAGAGTCTCAGATATCTGGGAGGAGGAGAGAGATCAGTCACCATCTTAGACTTCT 456  
Db 77 Asp-----SerGluSer 80  
QY 457 GAGGAAGATAAGGAAAAGAGAGGTTCTCTGTCAAAATCCAGAGTGCCTTCCGGGA 516  
Db 81 GluGluAspGluAspLeuGlu-----ValProValProSerArgPheAsnArg 96  
QY 517 CACATAGCCAGAGGAGGCAAGAAATGAAAACAATAGTCTTCAAAATGAGGAAAAA 576  
Db 97 ArgValSer-----ValCysAlaGluThrTyrAsnProAspGluGlu 110  
QY 577 GAGGAAAAAC 585  
Db 111 GluGluAsp 113

## RESULT 3

US-09-764-891-3819  
; Sequence 3819, Application US/09764891  
; Publication No. US20030077808A1

## GENERAL INFORMATION:

; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 3819  
; LENGTH: 103  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-891-3819

Alignment Scores:  
Pred. No.: 0.0162 Length: 103  
Score: 114.50 Matches: 28  
Percent Similarity: 50.00% Conservative: 7  
Best Local Similarity: 40.00% Mismatches: 30  
Query Match: 6.98% Indels: 5  
DB: 10 Gaps: 1

## AF334735 (1-954) x US-09-764-891-3819 (1-103)

QY 169 ATTCCACAGGATTTGGGAATCTTCTGAAGGGCTGACACGGCGAGATTTCTGAGAGCAA 228  
Db 30 ValProTyrGlyLeuLysThrLeuLeuGluGlyIleSerArgAlaValLeuLysThrAsn 49  
QY 229 CCGGACATATACCAAGCTTTTCAGAGCTTATTTTGAGAGCTTCTAGAGAAAGAGAG 288  
Db 50 ProSerAsnIleAsnGlnPheAlaAlaIlePheGlnGluLeuThrMetTyrArgGly 69  
QY 289 AAAACCACTTTGATCCAGCAGATGGGGGAGT-----AAGTAGAGAGAC 333  
Db 111 -----

Db 201 SerCysPheLeu-----ValPheLeuLeuSerCysLeuLeuAlaPheValLeu 216  
 QY 297 GTTGGTTCTCTCTCTCTCTAGAAAGGCTCTCAAAAATAGCGTGCTGCACAAAAGCTGTAT 238  
 Db 217 SerCysPheLeuAlaPhePhe-----LeuLeuPheLeuSerCysLeuLeuSerCysPhe 234  
 QY 237 ATTGTCGGTGTCTCTCTCAGAAATCTCGCG-----TGTCAAGCCCTTCAAGAAG 190  
 Db 235 LeuValPheLeuLeuSerCysLeuLeuAlaPheValLeuSerCysPheLeuPhe----- 252  
 QY 189 ATTCCCCAATCTTGTGGGAATTGGTAGTGGGTGGAGAAATGGAATCGACATCTTCTT 130  
 Db 253 -----PheLeuSerPheLeu-PheLeuSerPhePheLeuSerCysPheLeu 267  
 QY 129 GGTAAAGAACTGCTATGGAACTCTCTCTCCGAGCTGTGCCGCGCGTTCGGGTGT 70  
 Db 267 uPhePheHisSerPheLeuLeuSerPheLeuSerPhePheLeuProSer----- 283  
 QY 69 TTTTCTAGTTGCTGGGTAAACCGTTTTTCTT 39  
 Db 284 -PheLeuProSerPheLeuSerPhePheLeu 293

### RESULT 5

US-10-292-798-1840  
 ; Sequence 1840, Application US/10292798  
 ; Publication No. US20030235833A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SUWA, MAKIHO  
 ; APPLICANT: ASAI, KIYOSHI  
 ; APPLICANT: AKIYAMA, YUTAKA  
 ; APPLICANT: ABURATANI, HIROYUKI  
 ; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
 ; FILE REFERENCE: 084335/166  
 ; CURRENT APPLICATION NUMBER: US/10/292,798  
 ; CURRENT FILING DATE: 2002-11-13  
 ; PRIOR APPLICATION NUMBER: 10/017,161  
 ; PRIOR FILING DATE: 2001-12-18  
 ; PRIOR APPLICATION NUMBER: JP 2001-246789  
 ; PRIOR FILING DATE: 2001-06-18  
 ; NUMBER OF SEQ ID NOS: 2070  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1840  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-292-798-1840

Alignment Scores:	
Pred. No.:	Length: 298
Score:	Matches: 75
Percent Similarity:	Conservative: 33
Best Local Similarity:	Mismatches: 91
Query Match:	Indels: 72
DB:	Gaps: 13

AF334735 (1-954) x US-10-292-798-1840 (1-298)

QY 759 GTTTCATCGTATGCTTATTARACAATGTCATGACAGATTTTCACACAGTAACGAAT 700  
 Db 65 LeuSerLeuAlaCysLeuLeu-----AlaPhe 73  
 QY 699 GTTATTTCACCATCAAATCTTCTCTCAGGAAGAAATGACATTATAAGAAGGTG 640  
 Db 74 ValLeuSerCysPheProValPheLeuLeuSerPhePheLeuSerPheLeuSerCysLeu 93  
 QY 639 ATGGAT-----TTGGAATTATTTTCATGTTCTCGGAGGTAAACACAGTGTCTCACTT 586  
 Db 94 LeuSerCysLeuLeuAlaCysPheArgAlaPheLeu-----LeuSerCysPhe 109  
 QY 585 GTTTTCCTCTTTTCCTCATTTTGAAGACTATTGTTGTTTTCATTCTTTTTCCTCTCTCT 526  
 Db 110 LeuPheLeuSerPhe---PheLeuSerPheTyrcyPhePheLeuAlaCysPheLeuAla 128

QY 525 GGCTATGTGTCCTCCGAGGAGCTTGGATTGACAGCAGCAACCTCTCTCTTTCTT 466  
Db : : : : :  
129 CysLeuLeuAlaPheValLeuSerCys---PheProValPhePheLeuSerPhePheLeu 147  
QY 465 APTCTCTCAGAGAGTCTTAAGATGGTGACTGATCTC-----TTC 424  
Db : : : : :  
148 LeuPheLeuSer-----CysLeuLeuSerCysLeuLeuAlaPhe 160  
QY 423 CTCCTCTCCAGATATCTGAGACTCTCTTGTGTAGGATCATTCTCAGGTGGTCTTG 364  
Db : : : : :  
161 ValLeuSerCysPheLeuAlaPhePheLeuSerPheValSerPheLeuLeuAlaPheLeu 180  
QY 363 CTC-----CTC 358  
Db : : : : :  
181 LeuAspCysPheArgAlaPheLeuLeuSerCysPheLeuSerPhePheCysPhePheLeu 200  
QY 357 GAATGCATGATTGTATAGAGCGGTCTTCTACTCTTCTCCCTCCCTTCTGCTGGATCAA 298  
Db : : : : :  
201 SerCysPheLeu-----ValPheLeuLeuSerCysLeuLeuAlaPheValLeu 216  
QY 297 GTTGGTTCTCTCTTCTCTAGAGGCTCTCAAAATAGGCTGTGCAAAAGCTGGTAT 238  
Db : : : : :  
217 SerCysPheLeuAlaPhePhe-----LeuLeuPheLeuSerCysLeuLeuSerCysPhe 234  
QY 237 ATTGTCCGTTGCTCTCTCAGAACTCTCGCG-----TGTCCAGCTCTCAAGAAG 190  
Db : : : : :  
235 LeuValPheLeuLeuSerCysLeuLeuAlaPheValLeuLeuSerCysPheLeuPhe----- 252  
QY 189 APTCCCAAAATCTTGTGGAATTCGTTAGTGGTGTGGAGATGGAATCGACATCTTCTT 130  
Db : : : : :  
253 -----PheLeuSerPheLeu-PheLeuSerPhePheLeuSerCysPheLeu 267  
QY 129 GGTAGAACTGCTATGGAACCTCTCTCTCTCCGAGCTGTGGCGCGGTTCCGGTGT 70  
Db : : : : :  
267 uPhePheHisSerPheLeuLeuSerPheLeuSerPhePheLeuProSer----- 283  
QY 69 TTTTCTAGTGTCTGGTAACCGTTTCTT 39  
Db : : : : :  
284 -PheLeuProSerPheLeuSerPhePheLeu 293

## RESULT 6

US-10-437-963-148275  
; Sequence 148275, Application US/10437963  
; Publication No. US2004012343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 148275  
; LENGTH: 511  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_48722C.1.pep  
US-10-437-963-148275

## Alignment Scores:

Pred. No.:	0.1	Length:	511
Score:	108.50	Matches:	63
Percent Similarity:	39.67%	Conservative:	33
Best Local Similarity:	26.03%	Mismatches:	95
Query Match:	6.61%	Indels:	51

DB: 16 Gaps: 10  
AF334735 (1-954) x US-10-437-963-148275 (1-511)  
QY 3 GCCCTTCTCGCGCCGCTAGTTTTTTTTTTTAAAGAAAAACGGTTATCCAGCAAC 62  
Db : : : : :  
31 SerProSerArgSerProSerProArgAsnArgArgArgAspArgSerProSerPro 50  
QY 63 TAGAAAAACAAC-----CGAACCAGCGCCAGCAGCTCGGA 98  
Db : : : : :  
51 TyrArgSerArgAspArgSerProSerProTyrArgAspArgArgArgGlnTrpSer 70  
QY 99 -----GAGAAGGAGGTTCCATAGCGAGTTCTTACCAAGAGATGTGATCC 146  
Db : : : : :  
71 ProTyrHisArgAspArgGlyArgAspValGluArgGlnTrpAlaArgAspAsp--- 89  
QY 147 ATTCTCCAAACCCACTACCGAAT-----TCCACAAGATTTCGGAA 188  
Db : : : : :  
90 -----HisProAlaProArgArgGlyGlyAlaGlyAlaGlyAlaTrpSer 106  
QY 189 TCTTCTGAAAGGCTGACACCGGAGATCTGAGAGACGACCGGACATATACAGCTTT 248  
Db : : : : :  
107 AlaSer-----AspAspAspAspGluGlnLeuysGly----- 118  
QY 249 TGCAGCAGCTTATTTTGTAGAGCTCTTAGAGAAAGAGAGAAACC----- 294  
Db : : : : :  
119 ---LeuThr-TyrPheGluTyrArgValArgGluValArgGluValArgLysSerMetLy 137  
QY 295 -----AACTTTGATCCAGCAGATGGGGAGTAAAGTAGAAGACCGCTTCTTA 341  
Db : : : : :  
137 sArgCysAlaTrpAsnIleThrProSerProArgArgGluGlyGluAspGluAspTy 157  
QY 342 TAACAATCATCATTCGAGGAGCAAGAACACCTGAGAGAAAGTATCTCTAAACAAGA 401  
Db : : : : :  
157 rGlyTyrSerAspGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 177  
QY 402 GTCTCAGATATCTGGAGGAGGAGAGACATCAGTCACCATCTTAGACTCTCTGAGGA 461  
Db : : : : :  
177 aserSerAspLysSerGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 193  
QY 462 AGATAAGGAAAAAGAGAGTGTCTGCTGTCAAAATCCAAAGCTCTCTCGGGGACACAT 521  
Db : : : : :  
193 rAspSerGlyGluSerAsp-SerLeuSerAspSerSerLys-----SerAspAspThr 211  
QY 522 AGCCAGAGAGGAGCAAGAAATGAAACAAATAGTCTTCAAAATG-----AGGA 572  
Db : : : : :  
211 rGArgLysLysGlyArgGlySerHisArgSerSerLysArgSerArgHisArg 231  
QY 573 AAAAGAGGAAAAAAGTGGAGACACTGGTTTTTACCTCCAGGAAACATGAAAAATAATCA 632  
Db : : : : :  
231 rGArgHisSerSerAspThrGluGlyAspAspAsnSerLysAlaGluGluAspSerG 251  
QY 633 AA 634  
Db : :  
251 lu 251

## RESULT 7

US-10-092-947A-38  
; Sequence 38, Application US/10092947A  
; Publication No. US20030134353A1  
; GENERAL INFORMATION:  
; APPLICANT: WOLFF, Anne M  
; APPLICANT: APPEL, Karen F  
; APPLICANT: PETERSEN, Jesper F  
; APPLICANT: POULSEN, Ulla  
; APPLICANT: ARNAU, Jose  
; APPLICANT: JACOBSEN, Mette D  
; TITLE OF INVENTION: MUCOR RECOMBINANT GENE EXPRESSION  
; FILE REFERENCE: WOLFF-3  
; CURRENT APPLICATION NUMBER: US/10/092,947A  
; CURRENT FILING DATE: 2002-12-27  
; PRIOR APPLICATION NUMBER: US 60/274,650  
; PRIOR FILING DATE: 2001-03-12





Percent Similarity: 36.89% Conservative: 24  
 Best Local Similarity: 25.24% Mismatches: 72  
 Query Match: 6.34% Indels: 58  
 DB: 15 Gaps: 9

AF334735 (1-954) x US-10-424-599-237561 (1-580)

```

QY 64 AGAAAAACACCGGACCG-----GGCGGCACAGCTCGGAGAGA 102
Db 205 ArgLysAspGluGlyThrValArgLeuValAlaLysGlyGlyThrArgSerAspGly 224
QY 103 AAGGAGGTTCCATAGCAGCTTCTACCAAGAGATGTCATTCCTCCACACCCAC 162
Db 225 AsnGlu-----LysLeuLysAspMetGlyLeuAspValLysLysLeuAsp 239
QY 163 TACCGAATCCACAGGATTTGGGAATCTTCTGAAGGCTGACAGCGAGATTCGTAGA 222
Db 240 -----GlyArgGlyIleGlnAspGluValArgProIleGlyAsn 252
QY 223 GAGCAACCG-----GACAAATATACAGCT 246
Db 253 AlaAlaProValGlnAsnHisValGlyAsnPheHisProArgValAspGlyIleProLys 272
QY 247 TTTCGACGACGCTATTTTCGAGCGCTTCTAGAGAAAGAGAGAAACCAACTTTGATCCA 306
Db 273 LeuLeuGlyLysTyPheGluArgAsnLeuGluAla-----Thr 285
QY 307 GCAGATGGGGAGTAGTAGAGACCGCTTCTATAACAATCATGCTATCGAGAGCAA 366
Db 286 ValGluGlyLysGluLysValLys-----GluLysLys 296
QY 367 GAACACCTGAGAAAGTATCTTAAACAAGAGAGTCTCAGATATCTGGGAGGAGAA 426
Db 297 AspGluLysGluLysAlaLysGluLysThrGluGluArgLysValLysGluLysLysAsp 316
QY 427 GAGACATCATGTCACATCTTAGACTTCTCTGAGGAA-----GAT 465
Db 317 GluGlyLysGluLysValLysGluLysLysAspAspLysArgAspLysArgLysAsp 336
QY 466 AGGAAAAAGAGAGTGTCTGCTGCTCAAAATCCAGTCGCTTCGGGGACACATAGCC 525
Db 337 LysGluLysGluLysLysGlyHisGlyLysAspLysAspArgAspLysLysLysLys 356
QY 526 AGAGAGGAGCGCAAG-----AAAATGAAACAAATAGTCTTCAAAATGAGGAAAA 576
Db 357 GluGluLysAlaLysGluHisGlyGluLeuLysThrThrGlu---GlnAsnLysLeuLys 375
QY 577 GAGGAAAAACAGTGAGGA 594
Db 376 GluSerAsnLysValGly 381

```

# RESULT 10

```

US-10-114-270-178
; Sequence 178, Application US/10114270
; Publication No. US20040030110A1
; GENERAL INFORMATION:
; APPLICANT: Guo, Xiaojia
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Miller, Charles E.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Liu, Ziaohong
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Li, Li
; APPLICANT: Vernet, Corine
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Gorman, Linda
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Smithson, Glenna
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Gerlach, Valerie

```

```

; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Taupier Jr., Raymond J.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Ji, Weizhen
; APPLICANT: Anderson, David W.
; APPLICANT: Liete, Mario W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Stone, David J.
; APPLICANT: MacDougall, John R.
; APPLICANT: Rothenberg, Mark E.
; TITLE OF INVENTION: No. US20040030110A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-322C
; CURRENT APPLICATION NUMBER: US/10/114,270
; CURRENT FILING DATE: 2002-11-27
; PRIOR APPLICATION NUMBER: 60/281,086
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/281,136
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/281,863
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/281,906
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/282,020
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/282,930
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/282,934
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/283,512
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/283,710
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/284,234
; PRIOR FILING DATE: 2001-04-17
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 470
; SEQ ID NO 178
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-114-270-178

```

Alignment Scores:  
 Pred. No.: 0.321 Length: 532  
 Score: 103.50 Matches: 41  
 Percent Similarity: 42.86% Conservative: 25  
 Best Local Similarity: 26.62% Mismatches: 59  
 Query Match: 6.31% Indels: 29  
 DB: 15 Gaps: 7

AF334735 (1-954) x US-10-114-270-178 (1-532)

```

QY 181 TTTGGGATCTTCTTGAGGGCTGACAGCGGAGATTCTGAGAGACACCGACAATATA 240
Db 164 PheGly-----GluSerLeuLeuSerAspMetLeuAspPheProAspThrLeu 180
QY 241 CCAGCTTTTGACGAGCCTATTTTGAG-----AGCCTT 273
Db 181 ProSerProGluAlaLeuLysPheLysIleLeuValLysAsnLysLysIleGlyThrLeu 200
QY 274 CTAGAGAAAAGAGAGAAACCAACTTTGATCCA-----GCAGATGGGGAGT 321
Db 201 LysGluThrHisGluArgLysGlySerAspLysArgGlyLysValGluGluThrLeu 220
QY 322 AAGGTAGAAGACCGCTTCTTATAACAATCATGCATTCGAGGAGCAAGACCACTGAGAAA 381
Db 221 GluValAlaAspGly-----GluGluGluGluGluGluGlu 233
QY 382 AGTGATCTCTAAACAAGAGAGAGTCTCAGATATCTCGGAGAGGAGGAGCATCATGTCACC 441

```

Db 234 GluGluGluGluGluGluSerPhe-----LysGluSerGluValLeuGluSer 251  
Qy 442 ATCTTAGACTCTTCTGAGGAAGATAAGAAAGAGAGGTTGCTCTCTCAAAATCCAA 501  
Db 252 ValLeu---GlyAspAsnGlnAspLysGluThrGlyValLysGlySerLeuLysIleAla 270  
Qy 502 GCTGCTTCCGGGGACACATAGCCAGAGAGGAGGCAAGAAATGAAACA----- 552  
Db 271 LeuAlaLeuSerAspLeuValIleThrLysAlaGluLysPheLysSerPheGlnHis 290  
Qy 553 AATAGTCTTCARAAATCAGGAAGAAAGAGAAACAACTGAGGA 594  
Db 291 SerArgLeuThrGlnGlnPheAsnGluAsnSerIleGly 304  
RESULT 11  
US-10-425-114-64123  
; Sequence 64123, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 64123  
; LENGTH: 535  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3357-049-B12\_FLI.pep  
US-10-425-114-64123

Alignment Scores:  
Pred. No.: 0.322 Length: 535  
Score: 103.50 Matches: 58  
Percent Similarity: 40.20% Conservative: 22  
Best Local Similarity: 29.15% Mismatches: 66  
Query Match: 6.31% Indels: 53  
DB: 15 Gaps: 9

AF334735 (1-954) x US-10-425-114-64123 (1-535)  
Qy 83 GCGCCACCGCTCGGAGAGAAAGGAGTTCCATAGGAGTCTTACCAAGAGATGCGA 142  
Db 10 AlaAlaProAlaArgArgTrpArgLeu-----LeuLeuAlaArgLeuCyPro 26  
Qy 143 TTCCATTCTCCACACCCACTACCGAATTC-----CACAAGATTGGGAATC----- 190  
Db 27 ProAlaPheProAspProSerProLeuLeuProHisGluAsnLeuGlyValSerSer 46  
Qy 191 -----TTCTTGAAGGCTGACACGCGAGATTCCTGAGAGAGCA----- 227  
Db 47 SerAspLysAlaValGlnAlaGluSerAlaCysPhe-GluGlyAlaTrpSerGlyAl 66  
Qy 228 -----ACCGGCAATATATACCAGCTTTTGCAGCAGCGCTATTATTGAGAGCC 271  
Db 66 aglyAlaMetGlyLysLysGlyLysTrpPheGlyAlaValLysLysValPheSerProGl 86  
Qy 272 TTCTAGAGAAAGAGAGAAACCACTTTGATCCAGCAAGATGGGAGTAAGGTAGAAG 331  
Db 86 uSerLysGluLysLysGluGluArgLeu-----ArgArgLysSerAlaGlySerAs 103  
Qy 332 ACCG-----CTTCTATAACAAT----- 348  
Db 103 nProThrProValAspLeuThrProSerThrSerLeu-GluValAsnLeuSerValProp 123

Qy 349 -----CATGCATTCTGAGAGCAAGAACCCACTGTGAGAAA 382  
Db 123 roProProAlaProProValLeuHisGlnAlaGluGluValGlyValProGlu---- 141  
Qy 393 GTGATCTTAACAGAGAGTCTCAGATATCTGGGAAGAGGAGAGACATCAGTCACCA 442  
Db 142 -----AlaGluGlnGlnSerLysHisValAlaValGluGluAlaProAlaAlaLap 160  
Qy 443 TCTTAGACTCTTCT-----GAGGAAGATAAGGAAAAAGAGAGGCTGTCTGTCAAAA 496  
Db 160 roAlaGlnAlaSerValLeuProProAlaValProThrGlnGluLeuAlaValLysI 180  
Qy 497 TCCAAGCTGCTTCGGGGACACATAGCCAGAGAGGCAAGAAATGAAA 549  
Db 180 leGlnThrAlaPheArgGlyTyLeuAlaArgAlaLeuArgAlaLeuArg 197  
RESULT 12  
US-10-101-487-51  
; Sequence 51, Application US/10101487  
; Publication No. US20020169125A1  
; GENERAL INFORMATION:  
; APPLICANT: LEUNG, DAVID W.  
; APPLICANT: BERGMAN, PHILIP A.  
; APPLICANT: LOFQUIST, ALAN  
; APPLICANT: PIETZ, GREGORY E.  
; APPLICANT: TOMPKINS, CHRISTOPHER K.  
; APPLICANT: WAGGONER JR., DAVID W.  
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES  
; FILE REFERENCE: 077319/0329  
; CURRENT APPLICATION NUMBER: US/10/101,487  
; CURRENT FILING DATE: 2002-03-20  
; PRIOR APPLICATION NUMBER: 60/277,705  
; PRIOR FILING DATE: 2001-03-21  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 51  
; LENGTH: 197  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic fusion  
; OTHER INFORMATION: protein  
US-10-101-487-51  
Alignment Scores:  
Pred. No.: 0.275 Length: 197  
Score: 103.00 Matches: 29  
Percent Similarity: 45.97% Conservative: 28  
Best Local Similarity: 23.39% Mismatches: 61  
Query Match: 6.28% Indels: 6  
DB: 13 Gaps: 1  
AF334735 (1-954) x US-10-101-487-51 (1-197)  
Qy 217 CTGAGAGAGCAACCGGACCAATATACCAGCTTTTTCAGCAGCGCTATTTTTCAGAGCCTTCTA 276  
Db 7 LeuTyLysMetProGluAsnLeu-----TyrPheGlnGlyGluGlu 20  
Qy 277 GAGAAAGAGAGAAACCAACTTTGATCCAGCAAGATGGGGAGTAAAGTAGAACCGCG 336  
Db 21 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 40  
Qy 337 TTCTATAACATCATGCTTCGAGGAGCAAGAACCCCTGAGAAAGTGATCCTTAACAA 396  
Db 41 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 60  
Qy 397 GAAGAGTCTCAGATATCTGGGAAGAGGAGAGACATCAGTCACCATCTTAGACTCTTCT 456  
Db 61 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 80  
Qy 457 GAGGAAGATAAGGAAAAAGAGAGGTTCTGCTGTCAAAATCCAAAGTCGCTTCCGGGA 516  
Db 457 GAGGAAGATAAGGAAAAAGAGAGGTTCTGCTGTCAAAATCCAAAGTCGCTTCCGGGA 516

Db 81 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 100  
QY 517 CACATAGCCAGAGAGAGCAAGAAATGAAACAAATAGTCTTCAAAATCAGCAAAA 576  
Db 101 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 120  
QY 577 GAGGAAACAAG 588  
Db 121 GluGluGluGlu 124

RESULT 13  
US-101-487-114  
; Sequence 114, Application US/10101487  
; Publication No. US20020169125A1  
; GENERAL INFORMATION:  
; APPLICANT: LEUNG, DAVID W.  
; APPLICANT: BERGMAN, PHILIP A.  
; APPLICANT: LOFOUIST, ALAN  
; APPLICANT: PIETZ, GREGORY E.  
; APPLICANT: TOMPKINS, CHRISTOPHER K.  
; APPLICANT: WAGNER JR., DAVID W.  
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES  
; FILE REFERENCE: 077319/0329  
; CURRENT APPLICATION NUMBER: US/101/487  
; PRIOR FILING DATE: 2002-03-20  
; PRIOR APPLICATION NUMBER: 60/277,705  
; PRIOR FILING DATE: 2001-03-21  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 114  
; LENGTH: 197  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  
US-101-487-114

Alignment Scores:  
Pred. No.: 0.275 Length: 197  
Score: 103.00 Matches: 29  
Percent Similarity: 45.97% Conservative: 28  
Best Local Similarity: 23.39% Mismatches: 61  
Query Match: 6.28% Indels: 6  
DB: 13 Gaps: 1

AF334735 (1-954) x US-101-487-114 (1-197)  
QY 217 CTGAGAGACCAACCGACAAATATACCAGCTTTTGACGAGCCTATTTTGAGAGCCTTCTA 276  
Db 7 LeuTyrLysMetProGlnAsnLeu-----Ty:PheGlnGlyGluGlu 20  
QY 277 GAGAAAGAGAGAAACCAACTTTGATCCAGCAGAAATGGGGAGTAAGGTAGAACCGC 336  
Db 21 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 40  
QY 337 TTCTATAACAATCATGCTATTCAGGAGCAAGAACCACTGAGAAAGATGATCCTTAACAA 396  
Db 41 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 60  
QY 397 GAAGAGTCTCAGATATCTCGGAGGAGGAGAGACATCAGTCACCATCTTAGACTCTTCT 456  
Db 61 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 80  
QY 457 GAGGAGATAGAGAAAGAGAGAGTGTCTGTCTGCTCAAAATCCAAAGTGCCTTCCGGGA 516  
Db 81 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 100  
QY 517 CACATAGCCAGAGAGCAAGAAATGAAACAAATAGTCTTCAAAATCAGCAAAA 576  
Db 101 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 120  
QY 577 GAGGAAACAAG 588

Db 121 GluGluGluGlu 124  
RESULT 14  
US-101-153-668-264  
; Sequence 264, Application US/10153668  
; Publication No. US20030092616A1  
; GENERAL INFORMATION:  
; APPLICANT: HONDA, Goichi  
; APPLICANT: MATSUDA, Akio  
; APPLICANT: MURAMATSU, Shuji  
; APPLICANT: ISHIZAWA, Kenya  
; TITLE OF INVENTION: STAT6 Activating Gene  
; FILE REFERENCE: 1254-0207P  
; CURRENT APPLICATION NUMBER: US/101/153,668  
; CURRENT FILING DATE: 2002-05-24  
; PRIOR APPLICATION NUMBER: US 60/293,172  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: US 60/316,031  
; PRIOR FILING DATE: 2001-08-31  
; PRIOR APPLICATION NUMBER: US 60/328,403  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: JP 2001-157043  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: JP 2001-260681  
; PRIOR FILING DATE: 2001-08-30  
; PRIOR APPLICATION NUMBER: JP 2001-313175  
; PRIOR FILING DATE: 2001-10-10  
; NUMBER OF SEQ ID NOS: 488  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 264  
; LENGTH: 238  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-101-153-668-264

Alignment Scores:  
Pred. No.: 0.29 Length: 238  
Score: 103.00 Matches: 29  
Percent Similarity: 48.75% Conservative: 10  
Best Local Similarity: 36.25% Mismatches: 26  
Query Match: 6.28% Indels: 15  
DB: 14 Gaps: 1

AF334735 (1-954) x US-101-153-668-264 (1-238)  
QY 351 TGCATTCCGAGGAGCAAGAACCCCTGAGAAAGTGATCTTAAACAAGAGTCTCAGAT 410  
Db 4 CysMetArgGlyThrLysGlnVal-GluLysAsnAspAspGlnLysLysGluGlnAs 23  
QY 411 ATCTGGGAGGAGGAGGAGACATCAGTCACCATCTTAGACTCTTCTGAGGAGATAAGGA 470  
Db 23 pGlyIleLysProGlu-----As 29  
QY 471 AAAAGAGAGAGTGTCTGTCTGCTCAAAATCCAAAGTCCCTCCGGGACACATAGCCAGAGA 530  
Db 29 pLysAlaHisLysAlaAlaThrLysIleGlnAlaSerPheArgGlyHisIleThrArgLys 49  
QY 531 GGAGCAAGAAAGAAATGAAACAAATAGTCTTCAAAATGAGCAAAAAGAGGAAACAAAG 588  
Db 49 sLysLeuLysGlyGluLysAspValGlnAlaAlaGluAlaAsnLys 68

RESULT 15  
US-09-778-927A-42  
; Sequence 42, Application US/09778927A  
; Patent No. US20020068342A1  
; GENERAL INFORMATION:  
; APPLICANT: KHOSRAVI, Rami et al.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID AND AMINO ACID SEQUENCES AND NOVEL  
; FILE REFERENCE: 2786-0160P  
; CURRENT APPLICATION NUMBER: US/09/778,927A  
; CURRENT FILING DATE: 2001-02-08

```
; PRIOR APPLICATION NUMBER: IL 134453
; PRIOR FILING DATE: 2000-02-09
; PRIOR APPLICATION NUMBER: IL135341
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 42
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (1)..(252)
; OTHER INFORMATION: Xaa = any amino acid, unknown, or other
US-09-778-927A-42

Alignment Scores:
Pred. No.: 0.294 Length: 252
Score: 103.00 Matches: 29
Percent Similarity: 48.75% Conservative: 10
Best Local Similarity: 36.25% Mismatches: 26
Query Match: 6.28% Indels: 15
DB: 9 Gaps: 1

AF334735 (1-954) x US-09-778-927A-42 (1-252)
QY 351 TGCATTCGAGGAGCAAGAACCTGAGAAAGTGCCTCAAAACAGAGAGAGTCTCAGAT 410
   |||||
Db 4 CysMetArgArgThrLysGlnVal-GluLysAsnAspAspGlnLysIleGluGlnAs 23
   |||||

QY 411 ATCTGGGAGGAGGAGAGACATCAGTCACCATCTTAGACTCTCTTCGAGGAAGATAAGGA 470
   |||||
Db 23 pGlyIleLysProGlu-----As 29
   |||||

QY 471 AAAAGAGAGGTTGCTGCTCAAAATCCAAAGTGCCTTCGGGGACACATAGCCAGAGA 530
   |||||
Db 29 physalaHisLysAlaalaThrLysIleGlnAlaSerPheArgGlyHisIleThrArgLys 49
   |||||

QY 531 GGAGGCAAGAAATGAAACAAATAGTCTTCAAAATGAGGAAAAAGGAAAAACAAG 588
   |||||
Db 49 sLysLeuLysGlyGluLysLysAspValGlnAlaAlaGluAlaGluAlaAsnLys 68
   |||||

RESULT 16
US-10-437-963-145629
; Sequence 145629, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 145629
; LENGTH: 582
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; LOCATION: (1)..(582)
; OTHER INFORMATION: Xaa = any amino acid, unknown, or other
US-09-778-927A-43

Alignment Scores:
Pred. No.: 0.415 Length: 582
Score: 102.50 Matches: 35
Percent Similarity: 45.08% Conservative: 20
DB: 1 Gaps: 1

US-10-437-963-145629
```

```
Best Local Similarity: 28.69% Mismatches: 38
Query Match: 6.25% Indels: 29
DB: 16 Gaps: 5

AF334735 (1-954) x US-10-437-963-145629 (1-582)
QY 253 GCAGCCTATTTTGAGAGCCTTCTA-----GAGAAAGAGAGAAACCAACTTTTCATCCA 306
   |||
Db 6 AlalystrPileLysSerValLeuGlyLysLysSerAlaLysSerAsn----- 22
   |||

QY 307 GCAGAATGGGGAGTAAAGTAGAC-----CGCTTCTATAACAATCATGCTCATTCGAG 360
   |||
Db 23 -----SerThrLysAlaLysAspLeuAlaLysAlaLysAsnLysProValLeu 39
   |||

QY 361 GAGCAAGAACCACTGAGAAAGTGCCTTAAACAGAGAGTCTCAGATATCTCGGAAG 420
   |||
Db 40 SerGluAspProValIleSerGluProAlaLeuValAsnSerHisAsnAspGlyAsn 59
   |||

QY 421 GAGNAGAGACATCAGTC----- 438
   |||
Db 60 AlaGluAsnCysLysLeuProAsnGlyValAlaValGluAlaMetGlyGlnGlyValGlu 79
   |||

QY 439 -----ACCATCTTAGACTCT-----TCTGAGGAAGATAAGGAAAAAGAGAGGTT 483
   |||
Db 80 AsnGlnAsnIleValGlySerLysAlaProThrSerProGluLysLeuSerGluGluLeu 99
   |||

QY 484 GCTGCTGTCAAAATCCAAAGTGCCTTCGGGGACACATAGCCAGAGAGGCAAGAA 543
   |||
Db 100 AlaAlaValLysAlaGlnAlaAlaPheArgGlyTyrlLeuAlaArgAlaPheArgAla 119
   |||

QY 544 ATGAAA 549
   |||
Db 120 LeuLys 121

RESULT 17
US-09-778-927A-43
; Sequence 43, Application US/09778927A
; Patent No. US20020068342A1
; GENERAL INFORMATION:
; APPLICANT: KHOSRAVI, Rami et al.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID AND AMINO ACID SEQUENCES AND NOVEL
; TITLE OF INVENTION: VARIANTS OF ALTERNATIVE SPLICING
; FILE REFERENCE: 2786-0160P
; CURRENT APPLICATION NUMBER: US/09/778,927A
; CURRENT FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: IL 134453
; PRIOR FILING DATE: 2000-02-09
; PRIOR APPLICATION NUMBER: IL135341
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(267)
; OTHER INFORMATION: Xaa = any amino acid, unknown, or other
US-09-778-927A-43

Alignment Scores:
Pred. No.: 0.377 Length: 267
Score: 102.00 Matches: 27
Percent Similarity: 49.30% Conservative: 8
Best Local Similarity: 38.03% Mismatches: 22
Query Match: 6.22% Indels: 14
DB: 9 Gaps: 1

AF334735 (1-954) x US-09-778-927A-43 (1-267)
QY 376 GAGAAAGTGCCTTAAACAGAGAGTCTCAGATATCTGGGAAGAGAGACATCA 435
   |||||
```

Db 41 GluLysAsnAspAspAspGlnLysIleGluGlnAspGlyIleLysProGlu----- 57  
QY 436 GTCCACATCTTAGACTCTTCTGAGGAGATAGAGAAAGAGAGAGGTTGCTGCTCAAA 495  
Db 58 -----AspLysAlaHisLysAlaAlaThrLys 66  
QY 496 ATCCAGCTTGCCTTCGGGGACACATAGCAGAGAGAGGCGCAAGAAATGAAAAAAT 555  
Db 67 IleGlnAlaSerPheArgGlyHisIleThrArgLysLysLeuLysGlyGluLysLysAsp 86  
QY 556 AGCTCTCAAAATGAGGAAAGAGGAAACAAAG 588  
Db 87 AspValGlnAlaAlaGluAlaGluAlaAsnLys 97  
RESULT 19  
US-10-156-761-9889  
; Sequence 9889, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIORITY FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIORITY FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIORITY FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 9889  
; LENGTH: 376  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-9889  
Alignment Scores:  
Pred. No.: 0.413 Length: 376  
Score: 102.00 Matches: 39  
Percent Similarity: 39.38% Conservative: 24  
Best Local Similarity: 24.38% Mismatches: 55  
Query Match: 6.22% Indels: 42  
DB: 14 Gaps: 5  
AF334735 (1-954) x US-10-156-761-9889 (1-376)  
QY 169 ATTCCACAGAGATTGGGAATCTTCTGAAGGCTGACACGGAGATTCTGAGAGAGCAA 228  
Db 58 MetProGln---PheAlaGlnLeuGlnLeuArgGlyGluValLeuAspAlaGly 76  
QY 229 CCGGACATATACACGCTTTGCACAGCCTATTTTGAGCCTCTA----- 276  
Db 77 ArgLysAlaValThrAlaAlaAlaAspArgGlyMetSerSerLeuAlaAspAlaLeuSer 96  
QY 277 -----GAGAAAGAGAG 288  
Db 97 AspArgThrAlaArgLeuGlyLysLysGlyGluGluGluGlyGluGluGluGly 116  
QY 289 AAACCAACTTTGATCCAGCAAGATGGGGAGTAGAGTAGAGACCGCTTCTATAACAAT 348  
Db 117 GluGluGluTyrAlaProAspGluGluGlyAlaGluGlyGluGlu----- 131  
QY 349 CATGCATTCGAGGACCAAGACCCTCGAAGAAAGTATCTTAACAAGAGAGCTCTCAG 408  
Db 132 GluGluGluGluGluGluGluProGluGluGluGluProGluGluGluGluGlu 151  
QY 409 ATATCTGGGAGGAGAGAGACATCAGTCACCATCTTAGACTCTTCTGAGGAGAGATAAG 468  
|||||  
|||||

Db 152 ProGluAlaAlaLysGluGlu-----GluGluGluGluGlu 163  
QY 469 GAAAGAGAGAGTTGCTGTGTCAAAATCCAAGCTGCTTCCGGGGACACATAGCCAGA 528  
Db 164 GluGluGluGluGlnPro-----GluGlyGluTyrGluGlu 175  
QY 529 GAGGAGGCAAGAAATGAAACAAATAGTCTTCAAAATGAGGAAAGAGGAAACAAAG 588  
Db 176 GluGluGluGluGluGluGluGluGluProGluGluGluGluGluGluGlu 195  
RESULT 19  
US-10-108-260A-4549  
; Sequence 4549, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA  
; FILE REFERENCE: HI-A0106  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; PRIORITY FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 4549  
; LENGTH: 390  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-108-260A-4549  
Alignment Scores:  
Pred. No.: 0.526 Length: 390  
Score: 101.00 Matches: 42  
Percent Similarity: 41.32% Conservative: 27  
Best Local Similarity: 25.15% Mismatches: 56  
Query Match: 6.15% Indels: 42  
DB: 15 Gaps: 8  
AF334735 (1-954) x US-10-108-260A-4549 (1-390)  
QY 181 TTTGGGAATCTTCTTGAAGGCTGACACGGGCTGATCTGAGAGACACCGACATATA 240  
Db 9 PheGly-----GluSerLeuLeuSerAspMetLeuAspPheProAspThrLeu 25  
QY 241 CCAGCTTTTGACGACGCTATTTTGAG-----AGCCTT 273  
Db 26 ProSerProGluAlaLeuLysPheLysIleLeuValLysAsnLysLysIleGlyThrLeu 45  
QY 274 CTAGAGAAAAGAGAGAAAACCACTTTTGATCCA-----GCAGAAATGGGGAGT 321  
Db 46 LysGluThrHisGluArgLysGlySerAspLysArgGlyLysValGluGluTyrGluGlu 65  
QY 322 AAGTAGAAGACCCCTTCTATAACATCATGCTTCAGAGAGCAAGAACCCCTGAGAAA 381  
Db 66 GluValAlaAspGly-----GluGluGluGluGluGluGlu 78  
QY 382 AGTCATCTTAAACAAGAGAGCTCTCAGATATCTGGGAGGAGGAGAGACATCAGTCACC 441  
Db 79 GluGluGluGluGluGluGluAspLysPhe-----LysGluSerGluValLeuGluSer 96  
QY 442 ATCTTAGACTCTTCTGAGGAGATAGGAA----- 471  
Db 97 ValLeu---GlyAspAsnGlnAspLysGluThrGlyValLysLysLeuProGlyValMet 115  
QY 472 -----AAAGAGAGGTTGCTGCTCAAAATCCAAGCTGCTTCCGGGGACACATA 522  
Db 116 LeuPheLysLysLysThrArgLysLeuLysIleAlaLeuAlaLeuSerAspLeuVal 135  
QY 523 GCCAGAGAGAGGCAAGAAAATGAAAACA-----AATAGCTTTCAAAATGAGAA 573  
Db 136 IleTyrThrLysAlaGluLysPheLysPheGlnHisSerArgLeuTyrGlnGlnPhe 155  
QY 574 AAAGAGGAAAACAGTCAGGA 594  
Db 156 AsnGluAsnAsnSerIleGly 162  
|||||  
|||||





QY 277 GAGAAAGAGAGAAACCACTTTGATCCAGCAGAGTGGGGAGTAAGTAGAACCGC 336  
Db 527 ---LysGlnGlnAlaAsnPhe-----TpaLaserLys----- 537  
QY 337 TTTCTATAACAATCATGCTTCGAGGAGCAAGAACCCCTGAGAAAAGTGATCTAAACAA 396  
Db 538 -----AlaValGluGluGlnAlaLysAlaLysSerGluProLeuLys 552  
QY 397 GAAGAGTCT-----CAGATATCTGGG 417  
Db 553 GluGluSerAlaValLysAsnAspArgThrSerLysProLysValLysLeuSerTyr 572  
QY 418 AAGGAGGAA-----GAGCATCATGTC 438  
Db 573 LysGluGlnArgGluLeuGluGlnLeuProGlnLeuLeuGluGluGluThrLysIle 592  
QY 439 ACCATCTTAGACTCTTCTGAGGAAGAT-----AAGGAAAAAGAGAGGTTCGT 486  
Db 593 ThrValLeuGlnAlaGluIleAlaAspProAlaPhePheGlnGlnAlaHisAspIleThr 612  
QY 487 GCTGTCAAAATCCAGTCCCTTCGCGGGACACATAGCCAGAGAGGAGGCAAGAAATG 546  
Db 613 AspAlaLysLeuLysAla-----LeuAlaAspThrGluAla---GluLeu 626  
QY 547 AAAACAAATAGTCTTCAAAATGAGGAAAGGAGGAGGAAACAG 588  
Db 627 GluThrAlaPheLeuArgTrpGluGluLeuGluGluLysLys 640

## RESULT 22

US-10-467-248-3

; Sequence 3, Application US/10467248

; Publication No. US20040086905A1

; GENERAL INFORMATION:

; APPLICANT: DAS, Debopriya; YAO, Monique G.;

; APPLICANT: ARVIZU, Chandra S.; BAUGHN, Mariah R.;

; APPLICANT: LU, Yan; HAPALIA, April J.A.;

; APPLICANT: CHAWLA, Navinder K.; GRIFFIN, Jennifer A.;

; APPLICANT: LU, Dying Aina M.; YUE, Henry;

; APPLICANT: DING, Li; ELLIOTT, Vicki S.;

; APPLICANT: FORSYTHE, Ian J.; RAMKUMAR, Jayalaxmi;

; APPLICANT: GANDHI, Ameena R.; ISON, Craig H.;

; APPLICANT: WARREN, Bridget A.; TANG, Y. Tom;

; APPLICANT: EMERLING, Brooke M.; HONCHELL, Cynthia D.;

; APPLICANT: LYNE, Michael; BARROSO, Ines;

; TITLE OF INVENTION: LIPID-ASSOCIATED MOLECULES

; FILE REFERENCE: PI-0358 USN

; CURRENT APPLICATION NUMBER: US/10/467,248

; CURRENT FILING DATE: 2003-08-06

; PRIOR APPLICATION NUMBER: PCT/US02/03813

; PRIOR FILING DATE: 2002-02-06

; PRIOR APPLICATION NUMBER: US 60/266,910

; PRIOR FILING DATE: 2001-02-06

; PRIOR APPLICATION NUMBER: US 60/276,891

; PRIOR FILING DATE: 2001-03-16

; PRIOR APPLICATION NUMBER: US 60/276,855

; PRIOR FILING DATE: 2001-03-16

; PRIOR APPLICATION NUMBER: US 60/279,760

; PRIOR FILING DATE: 2001-03-28

; PRIOR APPLICATION NUMBER: US 60/283,818

; PRIOR FILING DATE: 2001-04-13

; PRIOR APPLICATION NUMBER: US 60/285,405

; PRIOR FILING DATE: 2001-04-20

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PERL Program

; SEQ ID NO 3

; LENGTH: 649

; TYPE: PRT

; ORGANISM: Homo sapiens

; NAME/KEY: misc feature

; OTHER INFORMATION: Incyte ID No: 72852842CD1

US-10-467-248-3

Alignment Scores:  
Pred. No.: 0.604 Length: 649  
Score: 101.00 Matches: 42  
Percent Similarity: 41.32% Conservative: 27  
Best Local Similarity: 25.15% Mismatches: 56  
Query Match: 6.15% Indels: 42  
DB: 15 Gaps: 8

AF334735 (1-954) x US-10-467-248-3 (1-649)

QY 181 TTTGGGATCTTCTTGAAGGCTGCACGCGAGATTTCTGAGAGAGCAACCGACAATATA 240  
Db 268 PheGly-----GluSerLeuLeuSerAspMetLeuAspPheProAspThrLeu 284  
QY 241 CCAGCTTTTTCAGCAGCCTATTTTGAG-----AGCCTT 273  
Db 285 ProSerProGluAlaLeuLysPheLysIleLeuValLysAsnLysIleGlyThrLeu 304  
QY 274 CTAGAGAAAAGAGAGAAACCACTTTGATCCA-----GCAGATGGGGAGT 321  
Db 305 LysGluThrHisGluArgLysGlySerAspLysArgGlyLysValGluGluTrpGluGlu 324  
QY 322 AAGGTAGAGAGCGCTTCTATAACAATCATGCATTCGAGGAGCAAGAACCCACCTCAGAAA 381  
Db 325 GluValAlaAspGly-----GluGluGluGluGluGluGlu 337  
QY 382 AGTGATCTCTAAACAAGAGAGTCTCAGATATCTCGGAGAGGAGAGAGACATCAGTCACC 441  
Db 338 GluGluGluGluGluGluAspLysPhe-----LysGluSerGluValLeuGluSer 355  
QY 442 ATCTTAGACTCTCTGAGGAGAGATAAGGAA----- 471  
Db 356 ValLeu---GlyAspAsnGlnAspLysGluThrGlyValLysLysLeuProGlyValMet 374  
QY 472 -----AAAGAGAGAGTTGCTGCTCTCAAAATCCAAGTGCCTTCGCGGGACACATA 522  
Db 375 LeuPheLysLysLysLysThrArgLysLeuLysIleAlaLeuLeuSerAspLeuVal 394  
QY 523 GCCAGAGAGGAGCAAGAAATGAAACA-----AATAGTCTTCAAAATGAGGAA 573  
Db 395 IleTyrThrLysAlaGluLysPheLysSerPheGlnHisSerArgLeuTyrGlnGlnPhe 414  
QY 574 AAAGAGGAGAAACAAGTCAGGA 594  
Db 415 AsnGluAsnAsnSerIleGly 421

## RESULT 23

US-09-949-029-18

; Sequence 18, Application US/09949029

; Publication No. US20030134278A1

; GENERAL INFORMATION:

; APPLICANT: Karpen, G.H.

; APPLICANT: Dobie, K.W.

; APPLICANT: Kennedy, C.D.

; APPLICANT: Velasco, V.M.

; APPLICANT: McGrath, T.L.

; APPLICANT: Weko, J.

; APPLICANT: Patterson, R.W.

; TITLE OF INVENTION: Identification of chromosome inheritance modifiers in *Drosophila*

; FILE REFERENCE: 1211.015US1

; CURRENT APPLICATION NUMBER: US/09/949,029

; PRIOR FILING DATE: 2001-09-07

; PRIOR APPLICATION NUMBER: US 60/231,178

; PRIOR FILING DATE: 2000-09-07

; NUMBER OF SEQ ID NOS: 149

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 18

; LENGTH: 3067

; TYPE: PRT

; ORGANISM: *Drosophila melanogaster*

US-09-949-029-18



Db 1386 SerThrSerAlaProSerProGluLys 1394

RESULT 25

US-10-755-889-615

; Sequence 615, Application US/10755889

; Publication No. US20040171823A1

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB

; TITLE OF INVENTION: PATHWAY

; FILE REFERENCE: D0284 NP

; CURRENT APPLICATION NUMBER: US/10/755,889

; CURRENT FILING DATE: 2004-01-13

; PRIOR APPLICATION NUMBER: U.S. 60/440,068

; PRIOR FILING DATE: 2003-01-14

; PRIOR APPLICATION NUMBER: U.S. 60/469,757

; PRIOR FILING DATE: 2003-05-12

; NUMBER OF SEQ ID NOS: 823

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 615

; LENGTH: 2468

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-755-889-615

Alignment Scores:

Pred. No.:	0.976	Length:	2468
Score:	100.50	Matches:	45
Percent Similarity:	41.8%	Conservative:	32
Best Local Similarity:	24.4%	Mismatches:	73
Query Match:	6.12%	Indels:	34
DB:	16	Gaps:	5

AF334735 (1-954) x US-10-755-889-615 (1-2468)

QY 129 CAAGAAGATGTCGATTCCTCCACACACCCATCCGAAATTCACAAAGGATTGGGAA 188

Db 518 GlnLysAspLeuThrGlyGlnValProThrProValValLysGlnThrLysLeuLysGln 537

QY 189 TCTTCTTGAAGGCTGACACGCGAGATTCAGAGAGCAACCGGACAATATACCGACTTT 248

Db 538 -----ArgAlaAspSerArgGluSerLeuLysProAlaAlaLysProLeuProSer 554

QY 249 TGCACGACCTATTTTGAGCGCTTCTAGAGAAAGAGAGAAACCAACTTTGATCCAGC 308

Db 555 LysSer-ValArgLysGluSerLysGluGluThrProGluValThrLysValAsnHisVa 574

QY 309 AGAATGGGGGAGTAAAGTAGAGAC----- 333

Db 574 lGluLysProProLysValGluSerLysGluLysValMetValLysLysAspLysProVa 594

QY 334 -CGCTTCATACATCATGCAATTCGAGAGCAAGACACCTGAGAAAGTGATCCTAA 392

Db 594 llyserThrLysLysProSerValThrGluLysGluValProSerLysGluGluProSe 614

QY 393 ACAAGAAGTCTCAGATATCTGGAGAGAG----- 423

Db 614 rProValLysAlaGluValAlaGluLysGlnAlaThrAspValLysProLysAlaAlaLy 634

QY 424 -----GAAGACACATCAGTACCCTTCTAGACTCTTCTGAGGAA---GA 464

Db 634 sGluLysThrValLysLysGluThrLysValLysProGluAspLysLysGluLysGlu 654

QY 465 TAAGGAAAAGAGAGTTGCTGCTGCTCAAAATCCAAAGCTGCTTCCGGGACAC----- 519

Db 654 uLysProLysLysGluValAlaLysLysGluAspLysThrProLysLysGluGluLy 674

QY 520 -ATACCCAGAGGAGGCAAGAAATGNAACAATAGTCTTCAAAATGAGGAAAGA 578

Db 674 sProLysLysGluGluValLysLysGluValLysLysGluLysLysGluLysLy 694

QY 579 GGAAAAACAAG 588

Db 694 sGluProLys 697

RESULT 26

US-10-087-464-53

; Sequence 53, Application US/10087464

; Publication No. US20030059436A1

; GENERAL INFORMATION:

; APPLICANT: Chishti, Athar

; APPLICANT: Oh, Steven

; APPLICANT: Liu, David

; APPLICANT: Goel, Vikas

; APPLICANT: Li, Xuerong

; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof

; FILE REFERENCE: S1237/7019

; CURRENT APPLICATION NUMBER: US/10/087,464

; CURRENT FILING DATE: 2002-03-01

; PRIOR APPLICATION NUMBER: US 06/272,930

; PRIOR FILING DATE: 2001-03-02

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 53

; LENGTH: 743

; TYPE: PRT

; ORGANISM: Plasmodium falciparum

US-10-087-464-53

Alignment Scores:

Pred. No.:	0.887	Length:	743
Score:	99.50	Matches:	37
Percent Similarity:	41.06%	Conservative:	25
Best Local Similarity:	24.50%	Mismatches:	42
Query Match:	6.06%	Indels:	47
DB:	14	Gaps:	5

AF334735 (1-954) x US-10-087-464-53 (1-743)

QY 148 TTCTCCAAACCCACTACCGAATTCACAAAGGATT---GGGAATCTCTTGAAGG--- 201

Db 622 TyrThrAsnGlyAsnLysAsnIleGlnGlnIlePheGlnGlnAsnIleLeuGluAsnAsp 641

QY 202 -----CTGCACGCGAGATTCTGAGAGCAACCGGACAATATATACCGACTTTT 249

Db 642 ValLeuAsnGlnGluThrGluGluMetGluLysGlnValGluAlaIleThrLysGln 661

QY 250 GCACGACGCTATTTGAGAGCCTTCTAGAGAAAGAGAGAAACCAACTTTGATCCAGCA 309

Db 662 lIleGluAlaGluValAspAlaLeuAlaProLysAsnLysGlu----- 675

QY 310 GAATGGGGAGTAAAGTAGAGACCGCTTCTATAACAATCATGCTTCGAGGAGCAAGAA 369

Db 676 -----GluGluGluLys 679

QY 370 CCACCTGAGAAAGTGATCTCTAAACAAGAGAGTCTCAGATATCTGGGAGGAGAGAG 429

Db 680 GluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLysGlu 699

QY 430 ACATCAGTCACCTCTTAGACTCTTCTGAGAAAGTAAGAAAGAGAGAGGTTGCTGCT 489

Db 700 LysGlu-----GluLysGluLysGluLysGluLysGluLysGlu----- 712

QY 490 GTCAAAATCCAAGCTGCTTCCGGGGACACATACCGAGAGAGAGGAGGAGAAATGAAA 549

Db 713 -----LysGluGluGluLysGluLys 720

QY 550 ACAAAATAGTCTTCAAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 582

Db 721 GluGluGluGlnGluGluGluGluGluGluGluGluGlu 731

RESULT 27

US-10-282-122A-72013

; Sequence 72013, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:



```
QY 477 AGAGTTCCTGCTGTCACAAATCCAGTCTCCCTCCGGGACACATAGCCAGAGAGAGGC 536
Db 178 uGluProAlaAnThrProLysLysAlaAlaValHisAlaSerGluLysLysThrAlaTh 198
QY 537 AAAGAAATGAAACA---AATAGTCTTCAAAATGAGGAAAGAGGAAAC 595
Db 198 rlysgluProLysThrProThrGlySerGlnSerGluSerThrGluValAen 215

RESULT 29
US-10-363-616-339
; Sequence 339, Application US/10363616
; Publication No. US20040044181A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-113 (793)
; CURRENT APPLICATION NUMBER: US/10/363,616
; PRIOR FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: 09/654,935
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 490
; SEQ ID NO 339
; LENGTH: 670
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-363-616-339

Alignment Scores:
Pred. No.: 0.968 Length: 670
Score: 99.00 Matches: 71
Percent Similarity: 39.35% Conservative: 62
Best Local Similarity: 21.01% Mismatches: 120
Query Match: 6.03% Indels: 85
Db: 15 Gaps: 13

AF334735 (1-954) x US-10-363-616-339 (1-670)
QY 66 AAAAACAACCGAAGCGGCGGACAGCTCGAGAGAAAGGAGTTCACATA----- 116
Db 70 LysAsnLysLysGlyGlyAlaArgLysLysGlyLysGlyLysGlyLysGlyLysGln 89
QY 117 -----GGCAGTCTTTACCAAGAAGATGTCGATTCATCTCCAAACACCCACTA 164
Db 90 AlaAsnLysThrAlaSerSerGlySerSerAspLysAspSerSerAlaGluSerAla 109
QY 165 CGAATTCACAGAGATTTGGAAATCTCTTGAAGCGGTGACACCGAGATTCGAGAGA 224
Db 110 ProGluGluGlyGluValSerAspSerAspSerAspSerSerSerSerSerSer 129
QY 225 ---GCAACCGGACATATACAGCTTTTGCAGCAGCCTATTTTGAGAGCCTCTA---- 276
Db 130 AspSerSerSerGluAspGluGluPheHisAsp-GlyTyGlyGluAspLysMetGlyAs 149
QY 277 -----GAGAAAGAGAGAGAAACCAACTT 299
Db 149 pGluGluAspArgAlaArgLysGluGluMetThrGluLysGluArgGluGluLysPhe 169
QY 300 TGATCCAGCAGATGGGGAGTAAGGTAGAGACCGCTTCTATACATCATCATTCGA 359
Db 169 eAsnArgLysGlyLysGlyValLysGlyValLysGlyValLysGlyValLysGly 189
QY 360 GGAGCAAGAACCCACCTCGAAGAGGATCCCTAAACAAGAGAGTCTCAGATATCTGGAA 419
Db 189 sThrAlaLysLysGlyLysGlyLysGlyLysGlyLysGlyLysGlyLysGlyLys 209
QY 420 GGAGGAGAGACATCAGTCACCATCTTAGACTCTTCTGAGGAGAGATAAGGAA----- 471
Db 209 sLysLysLysLeuThrGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 229
QY 472 ----AAAGAAGAGTTCGTGCTGCTCAAAATCCAAATCCAAATCCAAATCCAAATCC 527
Db 229 rlysgAspGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 249
```

```
QY 528 AGAGGAG-----GCAAAGAAA-----ATGAAACAAA 554
Db 249 gGluysArgLysAsnArgThrAlaGluLeuAlaLysLysGlnProLeuLysThrSe 269
QY 555 TAGTCTTCAAAATGAGGAAAGAGGAAACAAAGTGGAGGAC----- 595
Db 269 rGluValTySerAspGluGluGluGluGlu-AspAspLysSerSerGluLysSera 289
QY 596 -----ACTGGTTTACCTCCAGGAAACATGAAATAATATCCAAA----- 634
Db 289 sArgSerSerArgThrSerSerSerAspGluGluGluGluGluGluGluProL 309
QY 635 -----TCCATCAACCTTCTTATTAATGTCATTTCTTCTCTGAGGAGGAAAGA 680
Db 309 ysSerGlnProValSerLeuProGluGluLeuAsnArgValArgLeuSerArgHisL 329
QY 681 TTTGATGTTGGAATAACATTCCTTACTGTTGTGGAATAATCTGTCATGAGCATTTGTTA 740
Db 329 eu-----GluArgTrpCysHisMetProPhePheA 339
QY 741 ATACCATACCATGAAACATGCCACTTGAAGATTTCTCTGAGATCATGAGTTTGTTCAC 800
Db 339 laLys---ThrValThrGlyCysPheValArgIleGlyIleGlyAsn-----H 354
QY 801 ACTTGTCTCAAGCCTATCTATAGAGACCTTTGGATTTAGAAATTATAGAACTAAAGTATCT 860
Db 354 isAsnSer-LysProValTyArg-----ValAla 363
QY 861 GAGATTACAGAGATCTCAGAGGTTATGTGTCTTAACATATATATCAATG 908
Db 364 GluIleThrGlyValValGluThr-----AlaLysValTyGlnLeu 377

RESULT 30
US-09-934-455-450
; Sequence 450, Application US/09934455
; Publication No. US20030121070A1
; GENERAL INFORMATION:
; APPLICANT: Adam, Luc
; APPLICANT: Creelman, Robert
; APPLICANT: Dubell, Arnold
; APPLICANT: Heard, Jacqueline
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddie, James
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Reuber, Lynne
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pineda, Omaiza
; TITLE OF INVENTION: Genes for Modifying Plant Traits IV
; FILE REFERENCE: MBI-0025
; CURRENT APPLICATION NUMBER: US/09/934,455
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/227439
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: MBI-0022
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: MBI-0023
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 516
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 450
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-934-455-450

Alignment Scores:
Pred. No.: 1.01 Length: 775
Score: 99.00 Matches: 45
Percent Similarity: 42.93% Conservative: 34
Best Local Similarity: 24.46% Mismatches: 54
Query Match: 6.03% Indels: 51
```

```
DB: 10 Gaps: 8
AF334735 (1-954) x US-09-934-455-450 (1-775)
QY 76 GGAACCGGGCCACAGCTCGAGAGAAAGAGAGTTCATAGGAGTTCCTACCAAGAG 135
Db 153 GlySerGlyArgAspThrAlaGlnGluGluGluPro----- 165
QY 136 ATGTGATTCATTCCTCCAAACCCACTACCAAGATTTCCCAAGGATTTGGG----- 186
Db 166 -----ArgThrProGluAsnValGlyLysSerAsn 175
QY 187 -----AATCTCTTGAAGGCTGACACCGGAGATTCCTGAGAGCAACCGGACATATA 240
Db 176 GlyArgLysArgLeuGluLysThrThrProGluLeuValSerAlaSerProAlaAsn--- 194
QY 241 CCAGCTTTTGCAGCAGCTTATTTGAGAGCTTCTAGAGAAAGAGAGAAACCACTTT 300
Db 195 ---SerMetAlaTrpAspTyrPhe-----PheMetValGluAsnMetProGlyProAsnLeu 212
QY 301 GATCCAGCAGAATGGGGAGTAGGTAGAGACCGCTTC-----TATAACAAAT 348
Db 213 AspAspArgGluValArgAsnGlyTyrGluAsnGlnSerSerHisPheGlnPheAsnGlu 232
QY 349 CATGATTCGAGAGCAAGAACACCTGAGAAAGTATCTTAACAAAGAGTCTCAG 408
Db 233 GluAspAspGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 252
QY 409 ATATCTGGGAG-----GAGGAGAGACATCAGTCACCATCTAGACTCT 453
Db 253 ---SerGlyLysValValGluGluMetGluProLysThrProGluLysValGluGluGlu 271
QY 454 TCTGAGGAATAGGAAAGAGAGTTCGCTGCTCAAAATCCAAAGCTGCTCCCG 513
Db 272 GluGluGluAspGluGluAspGlu----- 280
QY 514 GGACATAGCCAGAGAGAGGAGGAGAAAGAAATGAAATAGTCTTCAAAATGAGGAA 573
Db 281 -----GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 296
QY 574 AAGAGGAAA 583
Db 296 slyLysLys 599

RESULT 31
US-10-437-963-183719
; Sequence 183719, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 183719
; LENGTH: 1537
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_80783C.1.pap
US-10-437-963-183719
Alignment Scores: 1.21 Length: 1537
Pred. No.: 1537
```

```
Score: 99.00 Matches: 45
Percent Similarity: 39.89% Conservative: 26
Best Local Similarity: 25.28% Mismatches: 66
Query Match: 6.03% Indels: 41
DB: 16 Gaps: 7
AF334735 (1-954) x US-10-437-963-183719 (1-1537)
QY 60 AACTAGAAAACACACCGGACCGGCGGACAGCTCGAGAGAAAGAGGTTTCCATAGGC 119
Db 146 AsnHisGluGluAsnGlyGlyLysHisGluGluAsnGlyGlyLysGlnAsnGluAsnGly 165
QY 120 AGTTCTTACCAGAAGATGTGATTCATTTCCAAACACCCACTACCGAATTCACAAAG 179
Db 166 GluLysCysGluGluAsnGlyGlyLysProAlaGln-----IleSerAspAspGlnAla 183
QY 180 ATTTGGGAATCTTCTGAAGGCTGACACCGGAGATTCGAGAGACCAACCGGACATAT 239
Db 184 ValAsnGlyCysAsnGlySerAlaAspLysLysAspLysGluGluGluThr----- 200
QY 240 ACCAGCTTTTGCAGCAGCTATTTTGAGAGCCCTTCTAGAGAAAAGAGAGAAACCACTT 299
Db 201 -----GluLysAspAsnGluLys-GluLys----- 208
QY 300 TGATCCAGCAGAATGGGGAGTAGGTAGAGACCGCTTCTATAACAATCATGCATTCGA 359
Db 209 -----GluAspLys-----GluGluGluThrGlu 216
QY 360 GGAGCAAGAACCCCTGAGAAAAGTATCTTAACAAAGAGAGTCTCAGATATCTGGGAA 419
Db 216 uLysAspAsnGluLysGluLysGluAsp---LysGluGluGluThrLysLysAspAsnGlu 235
QY 420 GGAGGAAGAGACATCAGTCACCATCTTAGACTCTTCTGAGGAAGATAGGAAAAGAA-- 477
Db 235 uLysGluLysGluGlnLeuMetGlyThrAspGluLysGluLysGluLysGluLysGlu 255
QY 478 -GAGTTGCTGCTGTCAAAATCCAAAGCTGCTTCCGGGGACACATAGCCAGAGAGAGGC 536
Db 255 pGluAsnGluGluGluLysLeuGlu-----GluGluGlu 266
QY 537 AAGAAAATGAACAAATAGTCTTCAAAATGAGGAAAAGAGGAAAACAAAG 588
Db 266 uLysLysAspLysGluGluLysLeuGluGluLysGluLysGluLysGluLysGlu 283

RESULT 32
US-10-087-192-1197
; Sequence 1197, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1197
; LENGTH: 426
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-087-192-1197
Alignment Scores: 0.961 Length: 426
Pred. No.: 98.50 Matches: 41
Percent Similarity: 41.26% Mismatches: 18
Best Local Similarity: 28.67% Mismatches: 53
```

```

Query Match: 6.00% Indels: 31
DB: 13 Gaps: 7
AF334735 (1-954) x US-10-087-192-1197 (1-426)

QY 40 AGAAAAACGGTTACCCAGCACTAGAAAAACAACCGGAACCGCGGCACCGAGCTCGGAG 99
DB 4 ArgSerLeuGlySerArgAlaAlaArgAspArgThrAlaAspArgGly----- 19
QY 100 AGAAGAGGTTCCATAGGCA---GTTCTTACCAAGAAGATGTGATTCATTCCTCAAC 156
DB 20 ArgGlyA-gA-gProGlyAlaSerAlaCystrGlyA-gMetSerIle----- 35
QY 157 ACCCACTACCAATCCACAGGATTTGGGAATCTTCTGAAGGCTGACACGCGAGATT 216
DB 36 -----GluileProAlaGlyLeuThrGluLeuLeuGlnGlyPheThrValGluVal 52
QY 217 CTGAGAGAGCAACCGGACAATATACAGCTTTTGCAGCAGCGCTATTTTGAGAGCCTTCTA 276
DB 53 LeuArgHisGlnProAlaAspLeuLeuGluPheAlaLeuGlnHisPheThrArgLeuGln 72
QY 277 GAGAAAAGAGAGAAAAC-----AACTTTGATCCAGCAGAATGGGGAGT 321
DB 73 GlnGluAsnGluArgLysGlyAlaAlaArgPheGlyHisGluGlyArgThr-TripGly--- 91
QY 322 AAGGTAGAACCGCTTCTATACCAATCATGCATTCGAGGAGCAAGAACCCCTGAGAAA 381
DB 92 -----AspAlaGlyAlaAlaAlaGlyGlyGlyThrProSerLys 104
QY 382 -----AGTGATCTAAACAAGAGAGTCTCAGATATCTGGGAAGAGGAA 426
DB 105 GlyValAsnPheAlaGluGluProMetArgSerAspSerGlu---AsnGlyGluGluGlu 123
QY 427 GAGACATCA 435
DB 124 GluAlaAla 126

RESULT 33
US-10-424-599-166145
; Sequence 166145, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 166145
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(273)
; OTHER INFORMATION: unsure at all xaa locations
;
; OTHER INFORMATION: Clone ID: PAT_MRT3847_121044C.1.pep
US-10-424-599-166145

Alignment Scores:
Pred. No.: 0.955 Length: 273
Score: 98.00 Matches: 43
Percent Similarity: 38.15% Conservative: 23
Best Local Similarity: 24.85% Mismatches: 39
Query Match: 5.97% Indels: 68
DB: 15 Gaps: 10

AF334735 (1-954) x US-10-424-599-166145 (1-273)

```

```

QY 217 CTGAGAGAGCAA-----CCGGACAAATATACCAGCTTTT 249
DB 33 LeuArgGluGlnLeuGluGlnLeuAlaGluArgAlaProGluGlyLeuProArgIle 52
QY 250 GCAGCAGCGCTATTTTGTAGAGCGCTTCTAGAGAAAAGAGAGAAAACCAACTTTGTATCCAGCA 309
DB 53 SerLysAlaValLeuAsnAspTyrSerGluLysIleGluAlaIle----- 67
QY 310 GAATGGGGAGTAAGGTAGAGACGGCTTCTATACCAATCATGCATTCGAGGAGCAAGAA 369
DB 68 -----AlaSerLysLeuVal-----AsnHisVal-----SerAsp 77
QY 370 CCACCT-----GAGAAAAGTGATCTCTAAACAAGAGAGTCTCAGATA 411
DB 78 ProProValProLysLysAspPheGluArgAsnSerValGluGluAsnSerSerGluIle 97
QY 412 TCTGGGAAGGAGGAGACATCATGCATTCACCATCTTAGAC----- 450
DB 98 -----GluGluThrLysGlnIleLeuLeuSerSerGlyLeuArgArgPro 113
QY 451 -----TCTTCTGAGGAAGATAAG-----GAAAAAGAGAGGTTGCT 486
DB 114 ValProAlaSerSerThrGluAspArgAlaHisLysProAlaGluThrAspHisIleSer 133
QY 487 GCTGTCAAAATCCAAAGCTCGCTTCCGGGGACACATAGCCAGAGAGGAG----- 534
DB 134 ProValLysLeuAspAlaAlaAlaHisIle***LysHisArgLysLeuGlnGlu 153
QY 535 -----GCAAAGAAAATGAAA----- 549
DB 154 AspLeuThrAspGluMetValValLeuAlaLysGlnLeuLysGluSerSerLeuThrMet 173
QY 550 ACAAATAGTCTTCAAAATGAGGAAAAGAGAAAACAAAG 588
DB 174 SerGlnSerLeuGlnAsnThrGluLysAspSerSerArg 186

RESULT 34
US-10-424-599-208282
; Sequence 208282, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 208282
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_30103C.1.pep
US-10-424-599-208282

Alignment Scores:
Pred. No.: 1.11 Length: 478
Score: 98.00 Matches: 31
Percent Similarity: 41.73% Conservative: 22
Best Local Similarity: 24.41% Mismatches: 40
Query Match: 5.97% Indels: 34
DB: 15 Gaps: 4

AF334735 (1-954) x US-10-424-599-208282 (1-478)

QY 265 GAGAGCGCTTCTAGAGAAAAGAGAGAAAACCAACTTTGTATCCAGCAGAAATGGGGAGTAAG 324
DB 18 AspSerLysGluLysLysAspGlnLysSerSerLysSerLysLysTrpPheGlyLys 37

```





Db 721 ysgluGlulysProLysAspValProGlu---LysLysLysAlaGluSerProValLysG 740  
QY 344 ACAATCATGCTTCAG-----GAGCAAG 367  
Db 740 luGluAlaValAlaGluValThrlleThrLysSerValLysValHisLeuGluLysG 760  
QY 368 ACCACCTGAGAAAGTATCTAAACAGAGAGCTCTCAG-----A 409  
Db 760 luthrLysGluGluGluLysProLeuGlnGlnGluLysGluLysAlaGlyGlyG 780  
QY 410 TATCTGGGAAGGAGGAGAGACATCAGTCACCATCTTCTGAGGAAGATAAGG 469  
Db 780 luGlyGlySerGluGluGluGlySer-----AspLysGlyAlaLysGlySera 796  
QY 470 AAAAGAGAGAGTTCCTGCT-----GTCAAAATCCAAG 502  
Db 796 rgLysGluAspIleAlaValaenGlyGluValGluGlyLysGluValGlnGlu 816  
QY 503 CTGCTTCGGGGACACATAGCCAGAGGAGGCAAGAAATGAAACAAATAGTCTTC 562  
Db 816 hrLysGluLysGlySer---GlyArgGluGluGluLysGlyValThrAenGlyLeuA 835  
QY 563 AAAATGAGGAAAGAGGAGGAAACAAAGTGCAGGA 594  
Db 835 spLeuSerProAlaAspGluLysLysGlyGly 845

RESULT 37

US-10-424-599-174175  
; Sequence 174175, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 174175  
; LENGTH: 197  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_12829C.1.psp  
US-10-424-599-174175

Alignment Scores:  
Pred. No.: 0.981 Length: 197  
Score: 97.50 Matches: 45  
Percent Similarity: 38.89% Conservative: 32  
Best Local Similarity: 22.73% Mismatches: 68  
Query Match: 5.94% Indels: 54  
DB: 15 Gaps: 7

AF334735 (1-954) x US-10-424-599-174175 (1-197)

QY 32 TTTTAAAGAAAAA-----ACGTTTACCAGCAACTAGAAAAACAACCG 76  
Db 37 PheThrLysGluGlnLeuArgValGlnValThrSerThrProValLeuArgIleAenGly 56  
QY 77 GAACCGGGCCACAGCTCGGAGAGAAAGAGGTTCCATAGCAGTTC---TTACCAAGA 133  
Db 57 GluArgGlnIleValGluAsnLysArgArgPheSerArgGluPheSerIleProPro 76  
QY 134 AGATGT-----CGATTCCATTCTCCAAACCCACTACCAATTC 172  
Db 77 TyrCysAspThrAsnAspValSerAlaLysPheGluGlyGlyValLeuSerIleLysPhe 96  
QY 173 CACAAGGATTTCGGAATCTTCTTGAAGGGCTGACACGGGAGATTCTGAGAGAGCAACGG 232

Db 97 ProLys----- 98  
QY 233 ACAATATACAGCTTTTCAGCAGCCTATTTTTCAGAGCCTTCTAGAGAAAAGAGAGAAA 292  
Db 99 -----LeuIleThrProAlaAlaArgSerGlnProGlnProGlnGluAla 113  
QY 293 CCAACTTGTATCCAGCAGATGG---GGGAGTAAAGGTAGAGACCGCTTCTATAACAAT 348  
Db 114 ProThrMetProGlnLysGluProSerGlnGlnGlnAspGlnValHisLysGln 133  
QY 349 CATGCATTCGAGGAGCAAGAACCCACCTGAGAAAAAGTGCATCTTAAACAAGAA---GAGTCT 405  
Db 134 GluSerLeuGlnLysGluLysGluProIleThrSerAspGluLysGluLysThr 153  
QY 406 CAGATATCTCGGAAGGAGGAGAGACATCATCATCATCTTGTAGACTCTTCTGAGGAAGAT 465  
Db 154 GluGluSerSerProGlnLysGluAsnGluProIleSerAspAspLysGluLysAsp 173  
QY 466 AAGGAAAAGAGAGGTTGCTGCTGTCCTCAAAATCCAGCTGCTTCCGGGGACACATAGCC 525  
Db 174 AsnLysThrGluGluValVala----- 181  
QY 526 AGAGAGGAGGCAAAACAAATGAAACAAATAGTCTTCAAAATGAGGAAAAAGAG 579  
Db 182 -----GluLysLysValArgThrAenGlyValProGluThrAlaLysGlu 196

RESULT 38

US-10-437-963-202437  
; Sequence 202437, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 202437  
; LENGTH: 574  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_97717C.1.psp  
US-10-437-963-202437

Alignment Scores:  
Pred. No.: 1.31 Length: 574  
Score: 97.50 Matches: 26  
Percent Similarity: 57.53% Conservative: 16  
Best Local Similarity: 35.62% Mismatches: 28  
Query Match: 5.94% Indels: 3  
DB: 16 Gaps: 2

AF334735 (1-954) x US-10-437-963-202437 (1-574)

QY 340 TATAACAATCATGCTTCGAGAGCAAGAACCCACTGAGAAAAGTATCTTAAACAAGAA 399  
Db 62 HisAenAenGluAlaValGlnGluValGlyArgGlyGluAenSerSerLeuGlnGlyGlu 81  
QY 400 -----GAGTCTCAGATATCTGGGAGGAGGAGAGACATCAGTCACCATCTTAGACTCT 453  
Db 82 ValValValArgAspValSerGlnAspLeuGluLysGlnAenThrValValSerAspAla 101  
QY 454 TCTGAGGAA---GATAAGGAAAAAGAGAGGTTGCTGTCTGTCATAAATCCAAAGCTGCCTTC 510

Db 102 SerAspProGluArgLeuArgGluGluGlnAlaAlaValysAlaGlnAlaAlaPhe 121  
QY 511 CGGGGACACATACCCAGAGAGAGGAGGCAAGAAATGAAA 549  
Db 122 ArgGlyTyrLeuAlaArgAlaPheArgAlaLeuLys 134

## RESULT 39

US-10-437-963-158544  
; Sequence 158544, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 158544  
; LENGTH: 620  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_58008C.1.pap  
US-10-437-963-158544

Alignment Scores:  
Pred. No.: 1.34 Length: 620  
Score: 97.50 Matches: 51  
Percent Similarity: 40.85% Conservative: 36  
Best Local Similarity: 23.94% Mismatches: 74  
Query Match: 5.94% Indels: 52  
DB: 16 Gaps: 11

AF334735 (1-954) x US-10-437-963-158544 (1-620)

QY 75 CGGAACCGGGCGCAC-----CAGTCGGAGAGAAAGGAGGTTCCATAGGCAGCT 122  
Db 93 ArgAsnArgArgAlaTyrLeuAspArgLeuValArgGluGlyGluTyrPheSerGlu 112  
QY 123 TCTTACCAAGAGATGTCGATTCATTCCTCCACACCCCACTACCGAATTCACAGGATT 182  
Db 113 GluAlaMetArgGluArgGluProTyrLeuHisGluTyrLeuGly-----ArgPhe 130  
QY 183 TGGGAATCTCTTGAGGGCTGACCGGAGATTCGAGAGCAACCGGACATATATA-- 240  
Db 131 GlnAspProThrGlyArgAlaMetAlaArgProGlyGluArgTrpSer-GluThrLeuMe 150  
QY 241 ----CCAGCTTTTTCGACGACCTATTTCGAGACCTTCTAGACAAAGAGAGAGAAACCAA 296  
Db 150 tArgArgAlaGluGluAlaValleValGluLysIleArgGlyGluGlnIleArgGlu 170  
QY 297 CTTTGTATCCAGCAGAATGG---GGGAGTAGGTAGAGACCGGCTTCTATAACAATCATGC 353  
Db 170 yValAspProSerGluTrpValGlyGlyGlyAlaGluGlu-----Al 184  
QY 354 ATTTCGAGGACAGACACCTGAGAAAGTGTATCTTAACAGAGAGAGTCTCAGATATC 413  
Db 184 aMetGluGluGlnGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 204  
QY 414 TGGGACGAGGAGAG-----429  
Db 204 uSerGluGluGluGlnGlnGlyMetAspMetGluGluLysGlyArgGlyAlaAspAsnSerSe 224  
QY 430 -----ACATAGTCACC---ATCTTAGACTCT-----453

Db 224 rGluAsnProSerValThrGluValValAspSerAsnLysThrAlaGlyAlaSerLysGln 244  
QY 454 -----TCTGAGGAAGATAAGAAAAAGAA---CAGGTTGCTGCTGTCAAAATCCAAGC 503  
Db 244 nThrLeuSerAlaGluAspMetGluAspGlnLeuGluGlnPheThrSerLeuMetGlnGln 264  
QY 504 TGCCTTCGGGGACACATAGCCAGAGAGAGGCAAGAAAGAAATGAAACAAATAGTCTTCA 563  
Db 264 nLysPhe-----LeuSerGlyGluAspSerGluHisMetAspTyrSerArgIleAs 281  
QY 564 AAATCAGGAAAAGAGAGAAACAACTGAGGACACTGG 600  
Db 281 pAsnAspGluMetLeuAspAsp-----HisTrp 290

## RESULT 40

US-10-092-900A-20  
; Sequence 20, Application US/10092900A  
; Publication No. US20040043382A1  
; GENERAL INFORMATION:  
; APPLICANT: Padigar, Muralidhara  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Taupier Jr., Raymond J.  
; APPLICANT: Pena, Carol E.A.  
; APPLICANT: Li, Li  
; APPLICANT: Zerhusen, Bryan D.  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Ji, Weizhen  
; APPLICANT: Gorman, Linda  
; APPLICANT: Miller, Charles E.  
; APPLICANT: Kikuda, Ramesh  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Gangolli, Esha A.  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Guo, Xiaojia Sasha  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Casman, Stacie J.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Liu, Yi  
; APPLICANT: Anderson, David W.  
; APPLICANT: Spaderna, Steven K.  
; APPLICANT: Catterton, Elina  
; APPLICANT: Leite, Mario W.  
; APPLICANT: Zhong, Hainong  
; APPLICANT: Alsobrook, John P.  
; APPLICANT: Lepley, Denise M.  
; APPLICANT: Rieger, Daniel K.  
; APPLICANT: Burgess, Catherine E.  
; TITLE OF INVENTION: No. US20040043382A1el Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-290C  
; CURRENT APPLICATION NUMBER: US/10/092,900A  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: USSN 60/274,322  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: USSN 60/283,675  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: USSN 60/338,092  
; PRIOR FILING DATE: 2001-12-03  
; PRIOR APPLICATION NUMBER: USSN 60/274,281  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: USSN 60/274,191  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: USSN 60/325,681  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: USSN 60/304,354  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: USSN 60/279,995  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: USSN 60/294,899  
; PRIOR FILING DATE: 2001-05-31

```
; PRIOR APPLICATION NUMBER: USSN 60/287,424
; PRIOR FILING DATE: 2001-04-30
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 768
; SEQ ID NO 20
; LENGTH: 749
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-900A-20

Alignment Scores:
Pred. No.: 1.41 Length: 749
Score: 97.50 Matches: 50
Percent Similarity: 38.58% Conservative: 48
Best Local Similarity: 19.69% Mismatches: 63
Query Match: 5.94% Indels: 93
DB: 15 Gaps: 11

AF334735 (1-954) x US-10-092-900A-20 (1-749)
QY 70 ACAACCGGAACCGCGGACCC-----AGCTCGGAG 99
Db SerThrGlySerGlyAsnThrGluHisSerCysSerSerGlnLysGlnLysSerIleGln 203
QY 100 AGAAGGAGGTTCCATAGGCACTTTACCAAGAGAGATGTCGATTCCTCCACACC 159
Db HisArgGlnThrGlnSerAspLeuThrIleGluLysIleSerAlaLeuGluAsnSerLys 223
QY 160 CACTACCGAATTCACAGAGATTGGGAATCTTCTGAAGGCGTCACACGCGAGATT--- 216
Db 224 AsnSerAspLeuGluLysLysGluGlyArg---IleAspAspLeuLeuArgAlaIleCys 242
QY 217 ---CTGAGAGACAAACCGGACCAATATACAGCTTTTGGCAGCAGCCTATTTGAGAGC--- 270
Db 243 AspLeuArgArgGlnIleAspGlnGlnLysMetLeuGluLysTyrlLysGluArgLeu 262
QY 271 -----CTTCTAGAGAAA----- 282
Db 263 AsnArgCysValThrMetSerLysLysLeuLeuIleGluLysSerLysGlnGluLysMet 282
QY 283 ---AGAGAGAAAACC-----AACTTTGATCCAGCAAGAA 312
Db 283 AlaCysArgAspLysSerMetGlnAspArgLeuArgLeuGlyHisPheThrThrSerAsp 302
QY 313 TGGGGAGTAAGGTAGAGACCGCTTCTATAACAATCATGCAATTCGAGAG----- 363
Db 303 HisGlyAlaLysPheThrGluGlnTrpThrAspGlyTyrlAlaPheGlnAsnLeuIleLys 322
QY 363 ----- 363
Db 323 GlnGlnGluArgIleAsnSerGlnArgGluGluIleGluArgGlnArgLysMetLeuAla 342
QY 364 -----CAAGAACCACTCTGAGAAAAGTGCCTCAACCAAGAA 399
Db 343 LysArgLysProProAlaMetGlyGlnAlaProProAlaThrAsnGlnLysGlnTrp 362
QY 400 GAGTCTCAGATATCTGGGAAGGAGGAGAGACATCAGTCACCATCTTAGACTCTTCTGAG 459
Db 363 LysSerLysThrAsnGlyAlaGluAsnGluThr---LeuThrLeu-----Lys 377
QY 460 GAAGATAGGAAAGAGAGAGGTTGCTGCTGCTCAAAATCCAGCTGCCTTCGGGGAGAC 519
Db 378 GluTyrlHisGlnGlnGluGluIlePheLysLeuArgLeu-----GlyHis 392
QY 520 ATAGCCAGAGAGGAGCA-----AAGAAA 543
Db 393 LeuLysLysGluGluAlaGluIleGlnAlaGluLeuGluArgLeuGluArgValArgLys 412
QY 544 ATGAAAACAAATAGTCTTCAAAATGAGGAAAAAGAGAAAC 585
Db 413 LeuHisIleArgGluValLysArgIleHisAsnGluAspAsn 426
```

RESULT 41

```
US-10-369-493-6247
; Sequence 6247, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6247
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-6247

Alignment Scores:
Pred. No.: 1.41 Length: 484
Score: 97.00 Matches: 55
Percent Similarity: 41.36% Conservative: 24
Best Local Similarity: 28.80% Mismatches: 66
Query Match: 5.77% Indels: 46
DB: 15 Gaps: 8

AF334735 (1-954) x US-10-369-493-6247 (1-484)
QY 803 AGTGTAACAACATCATGATCTCAGAGAAATCTTCAAGTGCATGTTTCAATGTGTATCT 744
Db 275 SerGluAsnArgAlaThrIleGluGluIleGlnSerSer-AlaCys---AsnArgLeuLe 293
QY 743 TATTAACAATGCTCATGACAGATTTTCACACAGTAACCAAGTGTATTTCACACATC 684
Db 293 uLeuGlySerIleLeuLeuAsn-----Phe-ArgHisA 304
QY 683 AAATCTTCTCTCCTCAGGAAGAAATGACATTAATAAGAGGTGATGATTGGATTATT 624
Db 304 snLeuLeuLeuTrpLeuArgLeuLeuLeuLeuLeuLeuSerPheLeuSerLeuPheLeuP 324
QY 623 TTTCATGTTTCTCGAGGTAAACACAGTGCTCTCATCTGTTTCTCTTTTCTCTCTATT 564
Db 324 he--LeuLeuLeuGlyLeuLeuLeuSerSerLeuValLeuLeuSerSerValIleL 343
QY 563 TGAAGACTATTGTTGTTTCTTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 504
Db 343 euLeuLeuPheIleIleSerThrIleLeuLeuLeuLeu----- 355
QY 503 GCTTGGATTTTGACAGCAGCAACCTCTTCTTTTCTCTCTCTCTCTCTCTCTCTCTCT 444
Db 356 -----PhePheLeuPheLeuPheLeu----- 363
QY 443 ATGGTGAAGTGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 384
Db 364 -----LeuPheLeuLeuLeuPheLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 378
QY 383 CTTTCTCAGGTGGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 336
Db 378 heLeuLeuArgGlyPheLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 398
QY 335 CGGTCTTCTACTTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 288
Db 398 euLeuPhePheLeuAspTyrlArgLeuCysTrpMetHisGlnSerProArgLysArgLysL 418
QY 287 TCTCTTTTCTCTAGAGGCTCTCAAAATA 259
Db 418 euLysAlaLeuLeuLysCysLeuLysVal 427
```



```
Db 185 GluArgLysAsnValLysGlyIleArgAspAspIleGluGluGluAspGlnGluAala 204
QY 508 TTCGGGGACACATAGCCAGAGAGGAGGCAAGAAATGAAACAAATAGTCTTCAAAAT 567
Db 205 TyrPheArgTyrMetAlaGluAsn-----ProThrAlaGlyValValGln 219
QY 568 GAGGAAAAAGAGGAAAC 585
Db 220 GluGluGluGluAspAsn 225

RESULT 44
US-10-338-279-2
; Sequence 2, Application US/10338279
; Publication No. US20030113791A1
; GENERAL INFORMATION:
; APPLICANT: Kaufmann, Joerg
; TITLE OF INVENTION: CIP130 INHIBITS CELL CYCLE PROGRESSION
; FILE REFERENCE: PP-01513.104/200130.456D3
; CURRENT APPLICATION NUMBER: US/10/338,279
; CURRENT FILING DATE: 2003-01-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1261
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-338-279-2
Alignment Scores:
Pred. No.: 1.83 Length: 1261
Score: 97.00 Matches: 49
Percent Similarity: 40.2% Conservative: 34
Best Local Similarity: 23.7% Mismatches: 65
Query Match: 5.91% Indels: 58
DB: 14 Gaps: 12

AF334735 (1-954) x US-10-338-279-2 (1-1261)
QY 49 GGTATCCAGCACTAGAAAAACACCGAACCGCGCG-----ACGAGTCGGAGAGA 102
Db 51 GlyGlyProGlyThrLysArgGlyPheGlyPheAlaIleSerAlaGlyLys 70
QY 103 AAGGAGGTTCCATAGCAGTTCTTACCAGAGAGTGCATTCATTCTCCACACCCAC 162
Db 71 LysGluGluPro-----LysLeuProGlnSerHisSerAlaPheGlyAlaThr--- 87
QY 163 TACCGAATTCACAGGATTTGGGAATCTTCTTGAAGGCGTCACACGGAGATTCGAGA 222
Db 88 ---SerSerSerGlyPheGly-----Lys 95
QY 223 GAGCAACCGGACAAATATACCACTTTT----- 249
Db 96 SerAlaProGlnLeuProSerPheTyrLysIleGlySerLysArgAlaAsnPheAsp 115
QY 250 ---GCAGAGCCTATTTGAGAGCCTTCTAGAGAAAGAGAGAAACCACTTGCAT--- 303
Db 116 GluGluAsnAlaTyrPheGluAsp-----GluGluGluAspSerSerAsnValAspLeu 133
QY 304 -----CCAGCAGATGGGGAGTAAGGTAGAGACCGCTCTCTATAACAATCATGCA 354
Db 134 ProTyrIleProAlaGlu-----AsnSerProThrArgGlnGlnPheHisSerLys----- 150
QY 355 TTCGAGGAGCAAGAACCACTCGAGAAAAGTGTATCTCTAAACAAGAA-----GAGTCTCAG 408
Db 151 -----ProValAspSerAspSerAspAspProLeuGluAlaPhe 164
QY 409 ATATCTGGAGGAGGAGACATCATGTCACCATCTTAGACTCTTCTGAGGAGATTAAG 468
Db 165 MetAlaGluValGluAspGlnAlaAlaArgAspMetLysArgLeuGluGluLysAspLys 184
QY 469 GAAAAAGAGAGTTGCTGCTCTCAAA-----ATCCAAGCTGCC 507
```

```
Db 185 GluArgLysAsnValLysGlyIleArgAspAspIleGluGluGluAspGlnGluAala 204
QY 508 TTCGGGGACACATAGCCAGAGAGGAGGCAAGAAATGAAACAAATAGTCTTCAAAAT 567
Db 205 TyrPheArgTyrMetAlaGluAsn-----ProThrAlaGlyValValGln 219
QY 568 GAGGAAAAAGAGGAAAC 585
Db 220 GluGluGluGluAspAsn 225

RESULT 45
US-10-017-161-2198
; Sequence 2198, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2198
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (151)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (354)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (385)..(418)
; OTHER INFORMATION: Variable amino acid
US-10-017-161-2198
Alignment Scores:
Pred. No.: 1.54 Length: 438
Score: 96.50 Matches: 58
Percent Similarity: 37.12% Conservative: 27
Best Local Similarity: 25.33% Mismatches: 48
Query Match: 5.74% Indels: 96
DB: 14 Gaps: 12

AF334735 (1-954) x US-10-017-161-2198 (1-438)
QY 600 CCAGTGCTCTCA-----CTTGTTTCTCTCTTTTCTCTCTTTTGAAGACTATTGTTT 547
Db 200 ProLeuSerSerSerLeuLeuPheLeuLeuPhePheLeuHisSerLeuPhePhe 219
QY 546 CATTTT-----CTTTGCCTCTCTCTCTGGCTATGTGTCCCG 511
Db 220 SerPhePhePheAsnAsnHisSerProSerLeuProProLeuProProPheSerPro 239
QY 510 GAAGCGAGCTTGGATTTTGACAGCAGCAACCTCTTTCTTTTC----- 469
Db 240 SerPheLeuPhePhePhePhePhePhePhePheTyrPhePhePhePheHis 259
QY 468 -----CTTATCTCTCTCAGAGAGTCTAAGATGGTGAATGATCTCTTTC 424
Db 260 PheSerProLeuLeuPheIle-----PheLeuPheSerPhe 271
QY 423 -----CTCCCTCCAGATATCTGAGACTCTTCTTGTGTTTAGGATCATTCTTCAGGTGG 370
```







QY 337 TTCTATAACATCATCGATTTCGAGGAGCAAGAACACCTGAGAAAAGTATCTTAACAA 396  
Db 104 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 123  
QY 397 GAAGAGTCTCAGATATCTCGGAGGAGGAGAGACATCAGTCACCATCTTAGACTCTTCT 456  
Db 124 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 143  
QY 457 GAGGAAGATAAGAAAAAGAGGTTGCTGCTGTCCTCAAAATCCAAAGTCCTTCGCGGGA 516  
Db 144 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 163  
QY 517 CACATAGCCAGAGAGAGCAAGAAAATGAAACAAATAGTCTTCAAAATGAGGAAAA 576  
Db 164 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 183  
QY 577 GAGGAAAAACAGTGAGGACACTGG 600  
Db 184 GluGluGluCysThrThrHisTrp 191

## RESULT 50

US-10-273-334-34  
; Sequence 34, Application US/10273334  
; Publication No. US20030129631A1  
; GENERAL INFORMATION:  
; APPLICANT: Pasternack, Gary R.  
; APPLICANT: Kocheavar, Gerald J.  
; APPLICANT: Brody, Jonathan R.  
; APPLICANT: Kodkol, Shrihari S.  
; TITLE OF INVENTION: GENE FAMILY WITH TRANSFORMATION MODULATING ACTIVITY  
; FILE REFERENCE: 031787.0076  
; CURRENT APPLICATION NUMBER: US/10/273,334  
; CURRENT FILING DATE: 2002-10-18  
; PRIOR APPLICATION NUMBER: US/09/591,500  
; PRIOR FILING DATE: 2000-12-06  
; PRIOR APPLICATION NUMBER: PCT/US98/26433  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: US 60/069,677  
; PRIOR FILING DATE: 1997-12-11  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 34  
; LENGTH: 249  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-273-334-34

Alignment Scores:  
Pred. No.: 1.48 Length: 249  
Score: 96.00 Matches: 35  
Percent Similarity: 42.42% Conservative: 21  
Best Local Similarity: 26.52% Mismatches: 49  
Query Match: 5.85% Indels: 27  
DB: 14 Gaps: 4

AF334735 (1-954) x US-10-273-334-34 (1-249)

QY 193 CTTGAAGGCTGACACGCGAGATTCTGAGAGAGCAACCGACAATATACAGCTTTTGCA 252  
Db 145 LeuAspGlyCysAspArgAsp---LysGluAlaProAsp-----SerAspAla 160  
QY 253 GCAGCCTATTTTGAAGCCTTCTAGAGAAAAGAGAGAAACCAACTTTTGATCCAGAGAA 312  
Db 161 GluGlyTyrrValGluGlyLeuAspAspGluGluGluAspGluAsp----- 175  
QY 313 TGGGGGAGTAAAGTAGACCGCTTCTATACATCATGCTTCGAGGAGCAAGAACCA 372  
Db 176 -----GluGluIutyrAspGluAspAlaGlnValValAspGluGlu 190  
QY 373 CCTGAGAAAAGTGATCCTAAACAAGAGAGTCTCAGATATCTGGGAGGAGGAGAGACA 432  
Db 191 AspGluAspGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 210

QY 433 TCAGTCACCATCTTAGACTCTTCTGAGGAGAGATAAGGAAAAAGAGAGGTTGCTGCTGC 492  
Db 211 GluGluGlyTyrrAsnAspGlyGluValAspAspGluGluAspGluGlu----- 226  
QY 493 AAAATCCAAAGCTGCCTTCGCGGGACACATAGCCAGAGAGGAGGCAAGAAAATGAACA 552  
Db 227 -----GluLeuGlyGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 238  
QY 553 AATAGTCTTCAAAATGAGGAAAAAGAGGAAAAACA 586  
Db 239 GluAsnLeu-LysMetArgGluLysMetMetThr 249

Search completed: February 15, 2005, 10:28:54  
Job time : 174 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 15, 2005, 09:52:34 ; Search time 32.5 Seconds  
(without alignments)  
5648.664 Million cell updates/sec

Title: AF334735  
Perfect score: 1641  
Sequence: 1 TCGCCCTCTCGCCGCG.....AAAAAAAAAAAAAAAAAAAA 954

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-Q/cn2.1/USPTO.spool\_p/AF334735/runat.15022005.094943.10099/app.query.fasta.1.1095  
-DB=PIR79 -OFMT=fasta -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=50 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=AF334735 @CGN 1.1 63 @runat.15022005.094943.10099 -NPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 79:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	778	47.4	151	2	I38243
2	569	34.7	146	2	I46506
3	558	34.0	149	2	S49527
4	134	8.2	189	2	T29159
5	118.5	7.2	404	1	OKHU2R
6	115.5	7.0	400	1	OKG02R
7	115	7.0	532	2	T06029
8	113.5	6.9	443	2	T48593
9	110	6.7	103	2	T45769
10	108	6.6	403	1	A43435
11	108	6.6	1132	2	S37932
12	107.5	6.6	805	2	E70474
13	106	6.5	1178	2	S78475
14	105.5	6.4	1351	2	C71607

15	104	6.3	304	2	S44897	ZK1236.2 protein -
16	104	6.3	699	2	T21379	hypothetical prote
17	103	6.3	238	2	I52638	neuronal growth-re
18	102.5	6.2	204	2	S45705	triachin 2 - rabbi
19	102	6.2	213	2	J00075	neuromodulin - gol
20	101.5	6.2	1003	2	T24545	hypothetical prote
21	101	6.2	419	2	T18450	hypothetical prote
22	101	6.2	475	2	S52893	hypothetical prote
23	101	6.2	647	2	E64170	ABC-type transport
24	101	6.2	1658	2	S55101	hypothetical prote
25	100.5	6.1	491	2	C97267	hypothetical prote
26	100	6.1	205	2	C30491	hypothetical prote
27	100	6.1	762	2	G88436	protein T048.13 l
28	100	6.1	783	2	F84514	hypothetical prote
29	100	6.1	791	2	T24435	hypothetical prote
30	99.5	6.1	414	2	A39205	nuclear localizati
31	99.5	6.1	689	2	T34025	hypothetical prote
32	99.5	6.1	858	2	S15762	neurofilament trip
33	99	6.0	189	2	T02570	hypothetical prote
34	99	6.0	241	2	S44893	ZK1236.7 protein -
35	99	6.0	644	2	S55395	neurofilament prot
36	99	6.0	1192	2	A71623	probable secreted
37	98.5	6.0	213	2	S45706	triachin 3 - rabbi
38	98.5	6.0	418	2	T15827	hypothetical prote
39	98.5	6.0	706	2	A45990	junctional sarcopl
40	98	6.0	916	2	A27864	neurofilament trip
41	98	6.0	2453	2	S60254	nuclear receptor c
42	97.5	5.9	1094	2	S49313	protein kinase - s
43	97	5.9	407	1	E0BEQ3	immediate-early pr
44	97	5.8	484	2	F88924	protein R02C2.2 l
45	97	5.9	918	2	T38786	translation initiat

ALIGNMENTS

RESULT 1

I38243  
zona pellucida binding protein Sp17 [similarity] - human  
C:Species: Homo sapiens (man)  
C>Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 09-Jul-2004  
C:Accession: I38243; S52921  
R:Richardson, R.T.; Yamasaki, N.; O'Rand, M.G.  
Dev. Biol. 165, 688-701, 1994  
A:Title: Sequence of a rabbit sperm zona pellucida binding protein and localization dur  
A:Reference number: I38243; MUID:95046885; PMID:7525387  
A:Accession: I38243  
A>Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: mRNA  
A:Residues: 1-151 <RES>  
A:Cross-references: UNIPROT:Q15506; EMBL:Z48570; NID:G695580; PIDN:CAA88459.1; PID:G6955  
R:Lea, I.A.; Richardson, R.T.; Widgren, E.E.; O'Rand, M.G.  
submitted to the EMBL Data Library, March 1995  
A:Description: Human Sp17: a sperm-zona binding protein.  
A:Reference number: S52921  
A:Accession: S52921  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <LEA>  
C:Superfamily: sperm surface protein Sp17

Alignment Scores:  
Pred. No.: 4,568-64 Length: 151  
Score: 778.00 Matches: 151  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 47.41% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x I38243 (1-151)

QY 136 ATGTGATTCATTTCTCCAGCCACCTACCAAGATTCGAGATTTGGATCTTCTT 195

Db 1 MetSerIleProPheSerAsnThrHisTyArgIleProGlnGlyPheGlyAsnLeuLeu 20





```
Db 77 Asp-----SerGluSer 80
QY 457 GAGGAGATAGGAAAAAGAGAGGTCTGCTGCTCAAAATCCAGCTGCTCCGGGGA 516
Db 81 GluGluAspGluAspLeuGlu-----ValProValProSerArgPheAsnArg 96
QY 517 CACATAGCCAGAGAGAGGCAAGAAATGAAACAAATAGTCTTCAAAATGAGGAAAAA 576
Db 97 ArgValSer-----ValCysAlaGluThrTyrAsnProAspGluGlu 110
QY 577 GAGGAAAAAC 585
Db 111 GluGluAsp 113

RESULT 6
OKBO2R
protein kinase (EC 2.7.1.37), cAMP-dependent, type II-alpha regulatory chain - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 15-Oct-1982 #sequence_revision 15-Oct-1982 #text_change 09-Jul-2004
C:Accession: A00618; S17058
R:Takio, K.; Smith, S.B.; Krebs, E.G.; Walsh, K.A.; Titani, K.
Biochemistry 23, 4200-4206, 1984
A:Title: Amino acid sequence of the regulatory subunit of bovine type II adenosine cycli
A:Reference number: A00618; MUID:85023306; PMID:6386045
A:Accession: A00618
A:Molecule type: protein
A:Residues: 1-400 <TAK>
A:Cross-references: UNIPROT:P00515
R:Braun, R.K.; Vulliet, P.R.; Carbonaro-Hall, D.A.; Hall, F.L.
Arch. Biochem. Biophys. 289, 187-191, 1991
A:Title: Phosphorylation of RII subunit and attenuation of cAMP-dependent protein kinase
A:Reference number: S17058; MUID:91378531; PMID:1654846
A:Accession: S17058
A:Molecule type: protein
A:Residues: 155-166 <BRA>
C:Comment: The inactive form of the enzyme is composed of two regulatory chains and two
our cAMP molecules.
C:Comment: Four types of regulatory chains are found: I-alpha, I-beta, II-alpha, and II-
C:Comment: Type II regulatory chains are phosphorylated by the activated catalytic chain
C:Comment: Type II regulatory chains mediate membrane association by binding to anchorin
merization.
C:Superfamily: cAMP-dependent protein kinase regulatory chain; cAMP receptor protein cyc
C:Keywords: acetylated amino end; autophosphorylation; cAMP binding; duplication; hetero
F:1-134/Domain: protein interaction <DIM>
F:135-256/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CAP1>
F:257-389/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CAP2>
F:1/Modified site: acetylated amino end (Ser) #status experimental
F:48,211/Binding site: phosphate (Thr) (covalent) (by proline-directed kinase) #status p
F:95/Binding site: phosphate (Ser) (covalent) (by autophosphorylation) #status experimen
F:204,213/Binding site: cAMP (Glu, Arg) #status predicted
F:334,343/Binding site: cAMP (Glu, Arg) #status predicted

Alignment Scores:
Pred. No.: 0.0184 Length: 400
Score: 115.50 Matches: 46
Percent Similarity: 39.58% Conservative: 30
Best Local Similarity: 23.96% Mismatches: 61
Query Match: 7.04% Indels: 55
DB: 1 Gaps: 8

AF334735 (1-954) x OKBO2R (1-400)
QY 157 ACCGACTACCGAATTCACAGGATTTGGGAATCTTCTTGAAGGCTGACACGCGAGATT 216
Db 1 SerHis11eGlnIleProGlyLeuThrGluLeuLeuGlnGlyTyrThrValGluVal 20
QY 217 CTGAGAGAGCAACCGACATATACAGCTTTTGAGCAGCGCTATTTTGAGAGCGCTTCTA 276
Db 21 LeuArgGlnArgProProAspLeuValAspPheAlaValAspTyrPheThrArgLeuArg 40
QY 277 GAGAAAAAGAGAAAAACC-----AACTTT 300
Db 41 GluAlaArgSerArgAlaSerThrProProAlaAlaProProSerGlySerGlnAspPhe 60
```

```
QY 301 GATCCA-----GCAGATGGGGGAGCTAAGGTAGAAGAC----- 333
Db 61 AspProGlyAlaGlyLeuValAlaAspAlaValAlaAspSerGluSerGluAspGluGlu 80
QY 334 -----CGTTC----- 339
Db 81 AspLeuAspValProIleProGlyArgPheAspArgValSerValCysAlaGluThr 100
QY 340 TATTAACAATCATGATTCGAGGAGCAAGAACCACTGAGAAAAGTGATCTCTAAACAGAA 399
Db 101 TyrAsnProAspGluGluGluAspThrAspProArgValIleHisProLysThrAsp 120
QY 400 GAGTCTCAGATATCTGGGAGGAGGAGACATCAGTCACCATCTTAGACTCTTCTGAG 459
Db 121 GlnGlnArgCysArgLeuGlnGluAlaCysLysAspIleLeuLeuPhe----- 136
QY 460 GAAGATAAGGAAAAAGAGAGTGTCTGCTCTCAAAATCCAAAGCTGCTCCGGGGACAC 519
Db 137 LysAsnLeuAspProGluGlnLeuSerGlnVal---LeuAspAlaMetPhe----- 152
QY 520 ATAGCCAGAGAGAGGAGGCAAGAAATGAAACAATAGTCTTCAAAATGAGGAAAAAGAG 579
Db 153 -----GluArgThrValLysValAspGluHisValIleAspGlnGlyAspAspGly 169
QY 580 GAAACCAAGTCAGACGACTGGTGTTCCTCCAGGAA 615
Db 170 AspAsn-----PheTyrValIleGlu 176

RESULT 7
T06029
hypothetical protein T28119.100 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
C:Accession: T06029
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15484
A:Accession: T06029
A:Molecule type: DNA
A:Residues: 1-532 <BEV>
A:Cross-references: UNIPROT:Q9T069; EMBL:AL035709; GSPDB:GN000062; ATSP:T28119.100
A:Experimental source: cultivar Columbia; BAC clone T28119
C:Genetics:
A:Gene: ATSP:T28119.100
A:Map position: 4

Alignment Scores:
Pred. No.: 0.0213 Length: 532
Score: 115.00 Matches: 50
Percent Similarity: 41.36% Conservative: 29
Best Local Similarity: 26.18% Mismatches: 90
Query Match: 7.01% Indels: 22
DB: 2 Gaps: 6

AF334735 (1-954) x T06029 (1-532)
QY 70 ACAACCGGAACCGCGGCGCACCGCTCGAGAGAAAGAGGTTCCATAGGAGTCTTACC 129
Db 226 ThrThrGluAsnGlySerAspAspGlyGluGlnGlnThrLysSerGluLeuAspSer 245
QY 130 AAGAAGATGTCGATTCCTCCATCTCCAAACACCCACTACCGAATTCACAA-----GGA 180
Db 246 LysThrGlyGluLysGlyPheSerAspSerAsnGlyGluLeuProGluThrAsnLeuSer 265
QY 181 TTTGGGAATCTTCTTGAAGGGCTGACGCGAGATTCTTGAGAGAGCAACCGGACAATATA 240
Db 266 ThrSerAsnAlaThrGlu-----ThrThrGluSerSerGlySerAspGluSerGlySer 283
QY 241 CCAGCTTTTGACGAGCGCTATTTTGAGAGCGCTCTAGAGAAAGAGAGAAACCACTTT 300
Db 284 SerGlyLysSerThrGlyTyrGlnGlnThrLysAsnGluGluAspGluLysVal 303
```



```
QY 301 GATCCAGCAGATCGGGGAGTAAAGTAGAAGACCGCTTCTATAC----- 345
Db 304 GlnSerSerGluGluSerValLysValLysGluSerGlyLeuAsnGluLysAspAlaSer 323
QY 346 ---AATCATGATTCGAGGAGCAAGAACCAACCTGTGAGAAAAGTGATCTAAACAAGAAGAG 402
Db 324 SerSerGlnAspGluSerLysGluGluLysProGluArgLysLysGluGluSerSer 343
QY 403 TCTCAGATATCTGGGAAGGAGGAA-----GAGACATCAGTCACC 441
Db 344 SerGlnGluGluLysGluGluProGluLysArgLysGluLysGluAspSerSerSer 363
QY 442 ATCTTAGACTCTTCTCAGAGAATAGAGAAAAGAGAGTTGCTGCTCTCATAAATCCAA 501
Db 364 GlnGluGluSerLysGluGluProGluAsnLysGluLysGluAlaSerSerSerGln 383
QY 502 GCTGCCTCCGGGACACATACCCAGAGAGGAGGCAAGAAATGAACAATATGCTCTT 561
Db 384 Glu-----GluAsnGluLysGluThrGluLysGluLysGluGluSerSerSer 401
QY 562 CAA-----AATCAGGAAAAGAGGAGGAAAACAAG 588
Db 402 GlnGluGlyAsnGluAsnLysGluThrGluLys 412

RESULT 8
T48593
hypothetical protein T22N19.110 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C:Accession: T48593
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duisterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24490
A:Accession: T48593
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-443 <BEV>
A:Cross-references: UNIPROT:Q9LYR0; EMBL:AL163572
A:Experimental source: cultivar Columbia; BAC clone T22N19
C:Genetics:
A:Map position: 5
A:Introns: 23/3; 125/3; 196/3; 240/3
A>Note: T22N19.110

Alignment Scores:
Pred. No.: 0.0287 Length: 443
Score: 113.50 Matches: 34
Percent Similarity: 45.61% Conservative: 18
Best Local Similarity: 29.82% Mismatches: 35
Query Match: 6.92% Indels: 27
DB: Gaps: 3

AF334735 (1-954) x T48593 (1-443)
QY 277 GAGAAAGAGAGAGAAACCACTTTGATCCAGCAGATGGGGAGTAGTAGAGACCCG 336
Db 25 GluLysArgArgLysTrpThrPhe-----TlpLysLeuArgLysLysArg 40
QY 337 TCTTATACATCATCGATTCGAGGAGCAAGAACCACTGAGAAAGTAGTCCTTAACAA 396
Db 41 LeuProSerIleThrAlaProProGluHisArgThrSerHisGluSerHisGluGluGln 60
QY 397 GAAGACTCTCAGATATCTCGGAGGAGGAGGAGACATCATCACCATC-----TTA 447
Db 61 LysGluGluLysValSerAspValGlyGluLysSerGlnValSerCysSerArgGlnLeu 80
QY 448 GACTCTTCTCAGAGAACTAAG----- 468
Db 81 AspSerIleGluGluSerLysGlySerThrSerProGluThrAlaAspLeuValGln 100
QY 469 -----GAAAAGAGAGGTGCTGCTGCTCAAAATCCAAAGTCGCC 507
Db 101 TyrGlnMetPheLeuAsnArgGlnGluGluValLeuAlaAlaThrArgIleGlnThrAla 120
```

```
QY 508 TTCGGGGACACATAGCCAGAGAGGAGGCAAGAAAATGAAA 549
Db 121 PheArgGlyHisLeuAlaArgLysAlaLeuArgAlaLeuLys 134
```

## RESULT 9

T45769

hypothetical protein F26O13.20 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 09-Jul-2004

C:Accession: T45769

R:Delveny, M.; Berger, C.; Cooke, R.; Greillet, F.; Laudie, M.; Mewes, H.W.; Lemcke, K.;

submitted to the Protein Sequence Database, December 1999

A:Reference number: Z23013

A:Accession: T45769

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-103 &lt;DEL&gt;

A:Cross-references: UNIPROT:Q9SD11; EMBL:AL133452

A:Experimental source: cultivar Columbia; BAC clone F26O13

C:Genetics:

A:Map position: 3

A:Introns: 51/3

A&gt;Note: F26O13.20

## Alignment Scores:

Pred. No.: 0.0505 Length: 103

Score: 110.00 Matches: 21

Percent Similarity: 71.11% Conservative: 11

Best Local Similarity: 46.67% Mismatches: 13

Query Match: 6.70% Indels: 0

DB: Gaps: 0

AF334735 (1-954) x T45769 (1-103)

```
QY 418 AAGGAGGAGAGACATCATGATCCATCTTAGACTCTTCGAGGAGATAGGAAAAGAA 477
Db 17 ArgArgSerGlnSerArgIleThrIleArgSerSerAlaProGluThrThrArgGlu 36
```

```
QY 478 GAGGTTGCTGCTGTCATAAATCCAAAGTCCCTTCGGGGACACATAGCCAGAGGAGGCA 537
Db 37 GluIleAlaAlaValLysIleGlnAlaPhePheArgGlyHisLeuAlaArgAlaPhe 56
```

```
QY 538 AAGAAAATGAAAACA 552
Db 57 LysAlaLeuLysSer 61
```

## RESULT 10

A43435

protein kinase (EC 2.7.1.37), CAMP-dependent, type II regulatory chain - fungus (Blasto

C:Species: Blastocladiella emersonii

C:Date: 04-Mar-1993 #sequence\_revision 28-Jul-1995 #text\_change 09-Jul-2004

C:Accession: A43435; B43435

R:Marques, M.V.; Gomes, S.L.

J. Biol. Chem. 267, 17201-17207, 1992

A:Title: Cloning and structural analysis of the gene for the regulatory subunit of CAMP

A:Reference number: A43435; MUID:92381035; PMID:1512258

A:Accession: A43435

A:Molecule type: mRNA

A:Residues: 1-403 &lt;MAR&gt;

A:Cross-references: UNIPROT:P31320; GB:M81714; NID:g167191; PIDN:AAA33016.1; PID:g167191

A&gt;Note: sequence extracted from NCBI backbone (NCBIN:111926, NCBIP:111927)

A:Accession: B43435

A:Molecule type: DNA

A:Residues: 1-371, 'r', 373-403 &lt;MAR2&gt;

A:Cross-references: GB:M81713; NID:g167189; PIDN:AAA33015.1; PID:g167190

C:Comment: Type II regulatory chains are phosphorylated by the activated catalytic chain

C:Comment: Type II regulatory chains mediate membrane association by binding to anchori

merization.

C:Genetics:

A:Introns: 52/1

C:Complex: The inactive form of the enzyme is composed of two regulatory chains and two

our CAMP molecules.



A:Accession: E70474  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-805 <AQF>  
A:Cross-references: UNIPROT:O67825; GB:AE000769; NID:g2984262; PID:AAC07794.1; PID:g2984262  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: InfB  
C:Superfamily: translation initiation factor IF-2; translation elongation factor Tu homolog  
C:Keywords: GTP binding; nucleotide binding; P-loop  
F:309-419/Domain: translation elongation factor Tu homology <TU>  
F:315-322/Region: nucleotide-binding motif A (P-loop)  
F:416-419/Region: GTP-binding NKXD motif  
F:452-454/Region: GTP-binding SAK/L motif  
F:321,322,342,416,417,419,452/Binding site: Mg-GTP (Lys, Thr, Asn, Lys, Asp, Ser) #  
Alignment Scores:  
Pred. No.: 0.111 Length: 805  
Score: 107.50 Matches: 53  
Percent Similarity: 42.49% Conservative: 29  
Best Local Similarity: 27.46% Mismatches: 74  
Query Match: 6.55% Indels: 37  
DB: 2 Gaps: 8  
AF334735 (1-954) x E70474 (1-805)  
QY 73 ACCGGAACCGCGCCACCGAGTCTGGAGAGAGAGGTTCATAGCGAGTCTTACCAAG 132  
Db 4 ThrLysThrLysArgValSerAspValAlaLysGluLeu-----GlyValLysSerLys 21  
QY 133 AGATGTCGATTCATTCCTCCACACCCACTCCGGAATTCACCAAGG----- 180  
Db 22 GluIle---IleGluPheLeuHenGluTyrTyProArgProAspGlyLysProTrpLys 40  
QY 181 -----TTTGGGATCTT 192  
Db 41 AlaSerHisGlyLeuAspGluGlnAlaLeuGluMetIleTyrAspAlaPheGlyLys 60  
QY 193 CTTGAAGGGCTGACACCGCGAGATTCAGAGAGCAACCGACAATATACAGCTTTGCA 252  
Db 61 GluGluGluLysGluGluValValThrGluGln-----AlaGlnAla 75  
QY 253 GCAGCTATTTCAGAGCCTCTTAGAGAAAGAGAGAAACCAACTTCGACGAGCA 312  
Db 76 ProAlaGluValGlu-----GluLysLysGluGluLysGluLysGluValIle 92  
QY 313 TGGGGGAGTAAAGTAAAGACCGCTCTATACATCATCGATTTCGAGGAGCAAGACA 372  
Db 93 ValGluGluValValGluGluLys---LysProGluValIleValGluGluLys 111  
QY 373 CTGAGAAAGTGATCCTAAACAGAGAGAGTCTCAGATATCTGGGAAGGAGGAGACA 432  
Db 112 LysLysGluGluGluLysLysGluGluLysProLysSerValGluGluLeu 131  
QY 433 TCAGTCACCATCTTACATCTTCTGAGGAGATAGGAAAAAGAGAGTCTCTGCTGTC 492  
Db 132 IleLysGluIleLeuGluLysLysGluLysGluLysGluLysValGluLysGlu 151  
QY 493 AAAATCAAGCTGCCTTCGG---GGACACATAGCCAGAGAGGAGCAAGAAATGAA 549  
Db 152 ArgLysGluGluLysValArgValValGluValLysLysGluLysGluLys 171  
QY 550 ACAATAGTCTCAAAATCAGAAAGAGGAGGAAACAG 588  
Db 172 -----LysGluGluLysLysGluGluLys 180

RESULT 13

S78475  
mannosylphosphorylation protein MNN4 - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YKL200c; protein YKL201c  
C:Species: Saccharomyces cerevisiae  
C:Date: 11-Mar-1998 #sequence revision 17-Apr-1998 #text\_change 09-Jul-2004  
C:Accession: S78475; S38037; S38038

R:Odani, T.; Shimma, Y.; Yoshifumi, J.  
submitted to the EMBL Data Library, January 1996  
A:Description: Cloning and Analysis of the MNN4 Gene Required for Phosphorylation of N-1  
A:Reference number: S78475  
A:Accession: S78475  
A:Molecule type: DNA  
A:Residues: 1-1178 <ODA>  
A:Cross-references: UNIPROT:P36044; EMBL:D83006; NID:g1752735; PID:d1012343; PID:g1752735  
R:Maia e Silva, A.; Bossier, P.; Vilela, C.; Fernandes, L.; Soares, H.; Guerreiro, P.;  
submitted to the Protein Sequence Database, March 1994  
A:Reference number: S38024  
A:Accession: S38037  
A:Molecule type: DNA  
A:Residues: 121-249, 'AIQLQT', 255, 'MALLRD', 262-390, 'LRISSN', 398-515, 'LG' <MAI>  
A:Cross-references: EMBL:D28200; NID:g486355; PID:g486356  
A:Experimental source: strain S288C  
A:Note: this sequence has been revised in reference S78475  
A:Note: this was assumed to be protein YKL200c  
A:Accession: S38038  
A:Molecule type: DNA  
A:Residues: 640-1178 <MAW>  
A:Cross-references: EMBL:D28201; NID:g486357; PID:g486359; MIPS:YKL201c  
A:Experimental source: strain S288C  
A:Note: this sequence has been revised in reference S78475  
A:Note: this was assumed to be the complete sequence of protein YKL201c  
C:Genetics:  
A:Gene: SGD:MNN4; MNN4  
A:Cross-references: MIPS:YKL201c; SGD:S0001684  
A:Map position: 11L  
C:Function:  
A:Description: required for phosphorylation of N-linked oligosaccharides  
C:Keywords: transmembrane protein  
F:28-44/Domain: transmembrane #status predicted <TM>

Alignment Scores:  
Pred. No.: 0.161 Length: 1178  
Score: 106.00 Matches: 43  
Percent Similarity: 42.1% Conservative: 27  
Best Local Similarity: 25.90% Mismatches: 70  
Query Match: 6.46% Indels: 26  
DB: 2 Gaps: 6  
AF334735 (1-954) x S78475 (1-1178)

QY 133 AAGATGTCGATTCATTCCTCCAAACCCACCTACCGAATTCACAGGATTGGG---AAT 189  
Db 981 LysSerAsnLeuAsnPheSerSerAsnSerPheLeuAspLeuAsnSerTyrGlyLeuAsp 1000  
QY 190 CTTCTTGAAGGGCTGACACGGGAGATTCTGAGAGAG-----CAACCGGAC 234  
Db 1001 LeuPheAlaProThrLeuSerAspValAsnArgLysGlyIleGlnMetPheAspLysAsp 1020  
QY 235 AATATACAGCTTTTTCAGAGCGCTTATTTGAGAGCCTTCTAGAG-----AAAAGA 285  
Db 1021 ProfileIleValTyrGluAspTyrAlaTyrAlaLysLeuLeuGluGluArgLysArgAG 1040  
QY 286 GAGAAACCAACTTGTATCCAGCAAGATGGGGAGTAGAGTAGAGACCGCTTCTATAC 345  
Db 1041 GluLysLysLysLysGluGluGluLysLysLysLysGluGluGluLysLysLys 1060  
QY 346 AATCATGATTCGAGGAGCAAGAACCACTCAGAAAGTGATCCTTAAACAAGAGAGTCT 405  
Db 1061 LysGluGluGluGluLysLysLysLysGluGluGluLysLysLysLysGluGlu 1080  
QY 406 CAGATATCTGGGAAGGAGGAGAGACATCAGTCACCATCTTTAGACTCTTCTGAGGAAGAT 465  
Db 1081 GluLysLysLysLysGluGluGluGluLysLysLysLys-----GlnGluGluGluLys 1098  
QY 466 AAGGAAAGAGAGAGGTTCGCTGCTCAAAATCCAAAGCTTCCTCCGGGGACACATAGCC 525  
Db 1099 LysLysLysGluGluGluLysLysLys----- 1108



```

A;Accession: I52638
A;Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A;Residues: 1-238 <RES>
A;Cross-references: UNIPROT:P17677; GB:S66541; NID:g440922; PIDN:AAB28649.1; PID:g440922
R;Kosik, K.S.; Orcchio, L.D.; Bruns, G.A.P.; Benowitz, L.I.; MacDonald, G.P.; Cox, D.R.
Neuron 1, 127-132, 1988
A;Title: Human GAP-43: Its deduced amino acid sequence and chromosomal localization in "
A;Reference number: S06437; MUID:90166498; PMID:3272162
A;Accession: S06437
A:Molecule type: mRNA
A;Residues: 1-238 <ROS>
A;Cross-references: EMBL:M25667; NID:g182969; PIDN:AA52516.1; PID:g182970
R;Ng, S.C.; de la Monte, S.M.; Conboy, G.L.; Karns, L.R.; Fishman, M.C.
Neuron 1, 133-139, 1988
A;Title: Cloning of human GAP-43: growth association and ischemic resurgence.
A;Reference number: JH0179; MUID:90166499; PMID:3272163
A;Accession: JH0179
A;Status: preliminary
A:Molecule type: nucleic acid
A;Residues: 1-238 <NGA>
C;Genetics:
A;Gene: GDB:GAP43
A;Cross-references: GDB:119972; OMIM:162060
A;Map position: 3q13.1-3q13.2
A;Introns: 10/3; 210/1
C;Superfamily: histone H1
C;Keywords: lipoprotein; phosphoprotein; thiolester bond
F;3,4/Binding site: palmitate (Cys) (covalent) #status predicted

Alignment Scores:
Pred. No.:      0.251      Length:      238
Score:          103.00     Matches:    29
Percent Similarity: 48.75%   Conservative: 10
Best Local Similarity: 36.25%  Mismatches: 26
Query Match:      6.28%     Indels:    15
DB:               2        Gaps:       1

AF334735 (1-954) x I52638 (1-238)
Qy      351 TGCATTTCGAGGACGAAGACCCTGAGAAAAGTGATCTTAACAAGAAGAGTCTCAGAT 410
Db      4 CysMeCArGargThrLySglnVal-GluLySAmAAspAspGlnLySAlleGluGlnAs 23
Qy      411 ATCTCGGAGGAGGAGGAGACATCAGTCACCATCTTAGACTCTTCTGAGGAAGATAAGGA 470
Db      23 pGIYlLeLySProGU-----As 29
Qy      471 AAGAAGAGGTTCGTGCTGTCCAAATCCAAGTCGCTCCGGGGACACATGCCAGAGA 530
Db      29 pYSAlaHaLSyAlaAlaThrLySleGlnAlaserPheArGLyHisIleThrArgly 49
Qy      531 CGAGGCCAAGAAATCAACAATAGTCTTCAAATGAGGAAGAAAGAGGAAACAAG 588
Db      49 sLYSLeULySGlyGluLySLyASpAspValGlnAlaAlaGluAlaGluAlaAbnLys 68.

RESULT 18
S45705
triachin 2 - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 10-Dec-1994 #sequence_revision 19-Apr-1996 #text_change 11-Jan-2000
C;Accession: S45705
R;Peng, M.; Fan, H.; Kirley, T.L.; Caswell, A.H.; Schwartz, A.
PEBS Lett. 348, 17-20, 1994
A;Title: Structural diversity of triadin in skeletal muscle and evidence of its existen
A;Reference number: S45704; MUID:94298946; PMID:8026576
A;Accession: S45705
A;Status: preliminary
A:Molecule type: mRNA
A;Residues: 1-204 <PEN>
C;Superfamily: histone H1

Alignment Scores:

```



```
Db 801 ArgValLeuGluGluGluMetGluMetLysArgLysAsnGluGluAlaLysIleLysLeu 820
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 499 CAACTGCTTCGG-----GGACACATAGCAGGAGGAGGCAAGAA 543
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 821 GluAlaGluMetArgLysAlaGluGlnAlaGluIleuArgArgGluGluLysSer 840
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 544 ATGAAACAATAAGTCTTCAAAATGAGGAAAGAGAA---AACAAAGTGAGGACACTGG 600
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 841 ArgAlaLeuLysLysLeuGlnLysGluGluThrAsnLysMetGluGlnMetAsnAsnCys 860
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 601 TTTTACTCTCAGAAACATGAAATAATCCAAATCCAT---CAACTCTCTTATTAATGT 657
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 861 ThrPheLeuGlnAsnValProLeuPheArgHisLeuHisProSerGlySerTyr----- 878
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 658 CATTTCTCTCAGGAAGAAAGATTTGATGTTGTGTAATAAATTAACATT----- 702
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 879 -----SerGluArgMetLeuAspGlyLysValIleGlyLeuTyrTyr 892
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 703 ---CGTTACTGTTGTGAAATCTGTCATGAGCATTTGTTTAAATAAGCATACCATTTGAAC 759
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 893 SerGlyTyrTrpCysGln----- 898
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 760 ATGCCACTTGAGATTTCTGAGATCATGAGTTTCTTT 798
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 899 ---ProSerArgAspPheThrProIleLeuAlaGlnPhe 910
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 21
T18450
hypothetical protein C0570c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18450
R:Lawson, D.; Bowman, S.; Barrell, B.
A:Reference number: Z18937
A:Accession: T18450
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1419 <LAW>
A:Cross-references: UNIPROT:O77337; EMBL:AL008970; NID:el407852; PIDN:CAA15592.1
C:Genetics:
A:Map position: 3
A:Introns: 15/3; 169/1
A:Note: C0570c

Alignment Scores:
Pred. No.: 0.413 Length: 419
Score: 101.00 Matches: 51
Percent Similarity: 41.67% Conservative: 29
Best Local Similarity: 26.56% Mismatches: 71
Query Match: 6.15% Indels: 41
DB: 2 Gaps: 8

AF334735 (1-954) x T18450 (1-419)
QY 33 TTTTAAAGAAACCGTTTACCCAGCACTAGAAACAAACCGGACCGGCGCACAG 92
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152 PheIleLysSerSerLysAsnMetSerSerPheValThrLysThrGlnLysGluAsnAsn 171
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 93 CTC---GGAGAGAAAGGAGGTTCCATA-----GGCAGTTCTTACCAGAAAGATGTCGAT 143
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 LeuAsnGlyAsnLysAsnAlaAspValThrHisGlySerThrSerGlnAspArgGlnIle 191
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 144 TCCATTCTCCAAACCCACTACCGAATTCACAAGGATTTGGAAATCTCTTGAAGGCGT 203
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 AsnAsnGluAspHisLysAsnValAsnAsnThrGlnIleAsnGlu----- 206
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 204 GACACCGGAGATTTCTGAGAGAGCAACCGGACATATATACCGATTTTGCAGCGCTATT 263
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 207 AspIle-GlnIleIleHisIleGluGluAsnIle-----GluAsnGluHi 222
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 264 TGAGAGCCTTCTAGAGAAAGAGAGAAACCAACTTTGATCCAGCAGAAATGGGGAGTAA 323
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 222 eAspLysLeuGluGluThrGlnGluLysMetGluAspValAlaGlnLysAspGluGl 242
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 324 GGTAGAACAGCCGCTTCTATACATCATGCTATTCAGAGGAGCAAGAACCCCTTGAGAAAAG 383
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 242 hGlnIleAspAsnLeuGluAsnGluLysAspThrGluLysGluAlaAspGlyGlyGl 262
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 384 TGATCTCTAAACAAGAGAGTCTCAGATATCTGGGAAGGAGGAGACATCATGTCACCAT 443
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 262 uAspLysGluLysGlu-----GlyLysGluGluGlu----- 272
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 444 CTTAGACTCTTCTGAGGAGATAAGGAAAAAGAGAGGTTGCTGCTGTCACAAATCCAAGC 503
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 273 ----AspLysGluGluGluAspLysGluGluGluAsp----- 283
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 504 TGCTCTTCGGGGACACATAGCCAGAGGAGGCAAGAAAAATGAAAAACAATAGTCTTCA 563
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 284 -----LysLysGluGluAspLysLysGluGluCysLysGluGluTyr 297
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 564 AAATGAG-----GAAAAAGAGAAAAACAAG 588
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297 rlysgluAspCysLysAspLysGluGluAspLys 308
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 22
S52893
hypothetical protein YMR044w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR044w
C:Species: Saccharomyces cerevisiae
C:Date: 08-May-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C:Accession: S52893
R:Odell, C.; Bowman, S.
A:Reference number: S52885
A:Accession: S52893
A:Molecule type: DNA
A:Residues: 1-475 <ODE>
A:Cross-references: UNIPROT:Q04213; EMBL:Z48502; NID:g695715; PIDN:CAA88410.1;
C:Genetics:
A:Gene: SGD:IOC4; MIPS:YMR044w
A:Cross-references: SGD:S0004647
A:Map position: 138

Alignment Scores:
Pred. No.: 0.419 Length: 475
Score: 101.00 Matches: 42
Percent Similarity: 40.52% Conservative: 20
Best Local Similarity: 27.45% Mismatches: 51
Query Match: 6.15% Indels: 40
DB: 2 Gaps: 7

AF334735 (1-954) x S52893 (1-475)
QY 220 AGAGAGCAACCGGACAATATACAGCTTTTGAGGAGCCTATTTTGAGAGCCTCTTAGAG 279
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 146 ArgLysAsnLysLysLeuIleProAlaTyrGluMetAla----- 158
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 280 AAAAGAGAGAACCAACCACTTTTCATCCA-----GCAGATGGGGAGT-----AAG 324
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 MetArgGlyLysAsnGlyLeuAspIleTrpGluPheValGluTyrGlySerTyrGlyLys 178
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 325 GTAGAGAGCCGCTTCTTATACCAATCATGCTTCGAGGAGCAAGAACCA----- 372
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 ProAspGluGluTyrValGluGluGluGluGluGluGluGluGluGluGluGluGlu 198
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 373 -----CCTGAGAAAGTATCTCTAAACAAGAGAGTCTCAGATATCTGGGAGGAGAA 426
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 199 IleArgProThrArgSerSerArgGlnLysArgAlaSerGluThrGluLys 218
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 427 -----GAGACATCAGTCACCATC 444
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 SerGluGlyGlyAsnSerAsnLysArgLysArgValThrArgSerThrArgGlnGlnAla 238
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 445 TTAGACTCTTCTGAGGAGAGATAAGGAAAAAGAGAGGTTGCTGCTGCTCAAAATCCAAGCT 504
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```











```

Db      72 SerAspSerGluSerGluAla-GluThrLysLysGluSerLysAspSerSerSe 91
QY      339 CTATAACATCATGCTTCGAGGAGCAACAACCTGAGAAAGTGATCCTTAACAAGA 398
Db      91 rSerSerAspSerSerAspGluGluGluGluGluGluGluGluThrLysLysG1 111
QY      399 AGAGTCTCAGATATCTGGGAAGGAGGAGGAGAGACATCAGTCACCATCTTAGACTCTCTGA 458
Db      111 uGluSerLysGluSerSerSerSerSerSerSerSerSerSerSerSerSer 128
QY      459 GGAAGATAAGAAAAAGAGAGGTGCTGCTGCTCAAAATCCAAAGCTGCCCTCCGGGAGCA 518
Db      128 pSerGluSerGluLysGluSerAsnAspLysLysLysLysLysLysLysLys 144
QY      519 CATAGCCAGAGGAGGCAAGAAATGAATAAATAGTCTTCAA----AATGAGGAAAA 575
Db      144 uAspAlaGluGluGluGluGluGluGluGluGluGluSerSerAsnLysLysGlnLysAsnGluGlu 164
QY      576 AGAGGAA 582
Db      164 rGluGlu 166

RESULT 31
T34025
hypothetical protein C32F10.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 16-Aug-2004
R:Accession: T34025
R:Sammons, L.; Wohldmann, P.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid C32F10.
A:Reference number: Z21465
A:Accession: T34025
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-689 <SAM>
A:Cross-references: UNIPROT:O01683; EMBL:AF000195; PIDN:AAC24268.1; GSPDB:GN00019; CESP:
A:Experimental source: strain Bristol N2; clone C32F10
C:Genetics:
A:Map position: 1
A:Introns: 79/3; 292/2; 606/3
C:Superfamily: HMG box homology

Alignment Scores:
Pred. No.: 0.605 Length: 689
Score: 99.50 Matches: 49
Percent Similarity: 43.56% Conservative: 39
Best Local Similarity: 24.26% Mismatches: 62
Query Match: 6.06% Indels: 52
DB: 2 Gaps: 10

AF334735 (1-954) x T34025 (1-689)
QY      27 TTTTITTTTTTAAAGAAAACGGTTTACCAGCACTAGAAAACACACCGAACCGCGG 86
Db      417 PheAspTyrLeuAsnLysLysGluLysLysLysLysLysLysLysLysLysLys 430
QY      87 CACCAAGCTCGGAGAGAGAGGAGGTTCATAGGAGGTTCCTTACCAAGAGAGATGTCGATTCC 146
Db      431 HisArgIleAspAsnLysSerAlaGlyTyrGlySerSerAspGluAspAspIleAsp-Pr 450
QY      147 ATTCTCAACACCCACTACCGAATCCACAGAGATTTGGGAATCTCTTTGAAGGGCTGAC 206
Db      450 oTyrLysSerThr-----VallysAlaGluGly----- 459
QY      207 ACGCGAGATCTTCAGAGAGCAACCGCAATATACCGAGCTTTTGCAGCGAGCCTATTTTGA 266
Db      460 -----ArgGluGlnAspAsp-----SerAspG1 469
QY      267 GAGCCTTCTAGAG-----AAAAGAGAGAAAACCAACTTTGA 302

```

```

Db      469 uSerThrAspGluAspTyrAspLeuAspLysAspMetLysLysGlnLysAsnAspLysAs 489
QY      303 TCCACCAAAATGGGGAGTAAGTAGAAGACCGCTTCTATACCAATCATGATTCAGGAGA 362
Db      489 pSerSerGluGlySerGlySerGluProAspAspGluTyrAspSerGlySerGluLysAs 509
QY      363 GCAGAACCACTCGAGAAAAGTGATCCCTAAACAAGAGAGTCTCAGATATCTGGGAGAGA 422
Db      509 pAlaSerGlyThrGlyGluSerAspPro-----AspGluGluAsnLysGluProLysLys 527
QY      423 GGAAGAGACATCAGTCACCATCTTAGACTCTTCTGAGGAGAGATAGGAAAAAGAGAGGT 482
Db      527 sLysGluSer-----LysGluLysLysAsnLysArgGluLysLysG1 541
QY      483 TGCTGCTCTCAAAATCCAAAGCTGCTCCGGGAGACATACGACAGAG-----GAGGC 536
Db      541 uLysProValLysGluLysAlaValLysLysLysLysLysLysLysLysLysLysLys 561
QY      537 AAAGAAAATGAAACA-----AATAGTCTTCAAAATGAGGAAAAAGA 578
Db      561 oLysArgAlaThrThrAlaTyrIleIleTrrPheAsnAlaAsnArgAsnSerMetLysG1 581
QY      579 GGA 582
Db      581 uAsp 582

RESULT 32
S15762
neurofilament triplet M protein - chicken
C:Species: Gallus gallus (chicken)
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
C:Accession: S15762; S08061; A27040
R:Zopf, D.; Dineva, B.; Betz, H.; Gundelfinger, E.D.
Nucleic Acids Res. 18, 521-529, 1990
A:Title: Isolation of the chicken middle-molecular weight neurofilament (NF-M) gene and
A:Reference number: S15762; MUID:90174973; PMID:2105668
A:Accession: S15762
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-858 <ZOP>
A:Cross-references: UNIPROT:P16053; EMBL:X17102; NID:G63688; PIDN:CAA34958.1; PID:G63686
R:Zopf, D.; Dineva, B.; Betz, H.; Gundelfinger, E.
submitted to the EMBL Data Library, November 1989
A:Reference number: S08061
A:Accession: S08061
A:Molecule type: DNA
A:Residues: 1-355,368-858 <ZOP>
A:Cross-references: EMBL:X17102
R:Zopf, D.; Hermans-Borgmeyer, I.; Gundelfinger, E.D.; Betz, H.
Genes Dev. 1, 699-708, 1987
A:Title: Identification of gene products expressed in the developing chick visual system
A:Reference number: A27040; MUID:88112814; PMID:3123320
A:Accession: A27040
A:Molecule type: mRNA
A:Residues: 350-546, 'R' 548-858 <ZOP>
A:Cross-references: GB:X05558; NID:G63685; PIDN:CAA29073.1; PID:G63686
C:Genetics:
A:Introns: 355/3; 397/2
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil

Alignment Scores:
Pred. No.: 0.622 Length: 858
Score: 99.50 Matches: 45
Percent Similarity: 36.27% Conservative: 25
Best Local Similarity: 23.32% Mismatches: 70
Query Match: 6.06% Indels: 53
DB: 2 Gaps: 6

AF334735 (1-954) x S15762 (1-858)
QY      121 GTTCTTACCAAGAGATGTCGATTCTCAATTCCTCAACACCCCACTACCGAATTCACAAGGA 180

```

Db 382 LeuLeuAsnValLysMetAlaLeuAsePilleGluLeuAlaAlaTyrArg----- 397  
 QY 181 TTTGGGAATCTTCTGAAGGGCTGCACACGCGAGATTCTGAGAGCAACCGGACAAATATA 240  
 |||||  
 Db 398 -----LysLeuLeuGluGly-----GluGluThrArgPhe 407  
 QY 241 CGAGCTTTTGCACAGCGCTATTTTGAGAGCTTCTAGAGAAAGA----- 285  
 |||||  
 Db 408 SerAlaPheSerGlySerIleThrGlyProIlePheThrHisArgGlnProSerValThr 427  
 |||||  
 QY 286 -----GAGAAACCACTTTGATCCAGCAGAAATGGGGAGTAAGSTA 327  
 |||||  
 Db 428 IleAlaSerThrLysIleGlnLysThrLysIleGluProProlLys-----LeuLysVal 445  
 |||||  
 QY 328 GAAGACCGCTTCATCAATCATGCTTCGAGAGCAAGCAACCACTGAGAAAAGTGAT 387  
 |||||  
 Db 446 GlnHisLysPheValGluGluIleIleGluGluThrLysValGluAspGluLysSerGlu 465  
 |||||  
 QY 388 CTTAAA-----CAAGAAGAGTCT 405  
 |||||  
 Db 466 MetGluAspAlaLeuSerAlaIleAlaGluGluMetAlaAlaLysAlaGlnGluGlu 485  
 |||||  
 QY 406 CAGATATCTGGGAGGAGGAGACATCAGTC----- 438  
 |||||  
 Db 486 GlnGluGluGluLysAlaGluGluGluAlaValGluGluAlaValSerGluLysAla 505  
 |||||  
 QY 439 -----ACCATCTTAGACTCTCTGAGGAAGATAAGGAAAAGAGAGGTGCTGCT 489  
 |||||  
 Db 506 AlaGluGlnAlaAlaGluGluGluLysGluGluGluAlaGluGluGluGluAla 525  
 |||||  
 QY 490 GTCAAATCCAGCTCCCTCCGGGACACATAGCCAGAGAGGAGCAAGAAATGAAA 549  
 |||||  
 Db 526 AlaLysSerAspAlaAlaGluGluGlyGlySerLysLysGluGluIleGluGluLysGlu 545  
 |||||  
 QY 550 ACAATAGTCTTCAAAATGAGGAAAAGAGGAAAACAG 588  
 |||||  
 Db 546 GluGlyGluGluAlaGluGluGluGluAlaGluAlaLys 558  
 |||||  
 RESULT 33  
 T02570  
 hypothetical protein At2g39320 [imported] - Arabidopsis thaliana  
 N;Alternate names: hypothetical protein T16B24.4  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 09-Jul-2004  
 C;Accession: T02570; H84815  
 R;Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,  
 submitted to the EMBL Data Library, August 1998  
 A;Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.  
 A;Reference number: Z14679  
 A;Accession: T02570  
 A;Status: translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-189 <R0U>  
 A;Cross-references: UNIPROT:O80949; EMBL:AC004697; NID:G3402671; PID:G3402675  
 A;Experimental source: cultivar Columbia  
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
 eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
 Nature 402, 761-768, 1999  
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A;Reference number: A84420; MUID:20083487; PMID:10617197  
 A;Accession: H84815  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-189 <STO>  
 A;Cross-references: GB:AE002093; NID:G3402675; PIDN:AAC28978.1; GSPDB:GN00139.  
 C;Genetics:  
 A;Gene: T16B24.4; At2g39320  
 A;Map position: 2  
 A;Introns: 9/3; 34/3; 108/1  
 Alignment Scores:  
 Pred. No.: 0.573 Length: 189

Score: 99.00 Matches: 31  
 Percent Similarity: 41.13% Conservative: 27  
 Best Local Similarity: 21.99% Mismatches: 43  
 Query Match: 6.03% Indels: 40  
 DB: 2 Gaps: 5  
 AF334735 (1-954) x T02570 (1-189)  
 QY 220 AGAGCAACCGGCAAAATATA-----CCAGCTTTTGCAGCAGCTTATTTGAGAGC 270  
 |||||  
 Db 80 GlnLysGluProAspLysValIleHisMetSerTyrLeuAlaGlyIleHisPheAsnSer 99  
 |||||  
 QY 271 CTTCTAGAGAAA-----ACAGAGAAAACCAACTTTGATCCAGCAGAA 312  
 |||||  
 Db 100 IleTyrLysLysAsnLysGluLysGlySerArgSerSerSerSerSerSerAlaVal 119  
 |||||  
 QY 313 TGGGGAGTAGTAGAGACCGCTTCTATACATCATGCATTCGAGAGAGCAGAACCA 372  
 |||||  
 Db 120 Trp---MetLysLeuGlnArgLysLysGluAsnGluAlaLysLysGluGluGlu 138  
 |||||  
 QY 373 CCTGAGAAAGTGATCCTAAACAAGAGAGTCTCAGATATCTGGGAAGGAGGAAGACA 432  
 |||||  
 Db 139 LysGluArgLysAspMetGluLysGluGluLysLys----- 150  
 |||||  
 QY 433 TCAGTCACCATCTTAGACTCTTCTGAGGAAGATAAGGAAAAGAGAGGTTCGCTGTC 492  
 |||||  
 Db 151 -----LysAspLysGluAspLysLysLysAsp----- 159  
 |||||  
 QY 493 AAAATCCAGCTGCTTCGCGGGACACATAGCCAGAGAGGAGGCAAGAAATGAAAACA 552  
 |||||  
 Db 160 -----LysGluAspLysLysLysLysLysLysLysLysLys 168  
 |||||  
 QY 553 AATAGTCTTCAAAATGAGGAAAAGAGGAAAACAAAGTAGGAGACACTGGTTTACCTCCAG 612  
 |||||  
 Db 169 GlnLysGluLysGluLysLysLysLysLysLysLysLysLysLysLysLysLys 188  
 |||||  
 QY 613 GAA 615  
 |||||  
 Db 189 Glu 189  
 RESULT 34  
 S44893  
 ZK1236.7 protein - Caenorhabditis elegans  
 C;Species: Caenorhabditis elegans  
 C;Date: 14-Sep-1994 #sequence\_revision 12-May-1995 #text\_change 23-Mar-2001  
 C;Accession: S44893  
 R;Favell, A.D.  
 submitted to the EMBL Data Library, May 1993  
 A;Description: Sequence of the C. elegans cosmid ZK1236.  
 A;Reference number: S44622  
 A;Accession: S44893  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-241 <PAV>  
 A;Cross-references: EMBL:L13200; NID:G289748; PID:G289749  
 C;Genetics:  
 A;Introns: 122/3; 209/1  
 Alignment Scores:  
 Pred. No.: 0.591 Length: 241  
 Score: 99.00 Matches: 49  
 Percent Similarity: 40.93% Conservative: 39  
 Best Local Similarity: 22.73% Mismatches: 94  
 Query Match: 6.03% Indels: 33  
 DB: 2 Gaps: 7  
 AF334735 (1-954) x S44893 (1-241)  
 QY 75 CGGAACCGCGCACCAGCTCGGAGAGAAAGAGGTTCCATAGGCAGTTCCTACCAAGAA 134  
 |||||  
 Db 3 ArgAsnAlaArgArgValAsnArg-----AapGlu 13  
 |||||  
 QY 135 GATGTCGATTCTCCACACCCACTACCGAATTCACCAAGGATTTGGGAATCTTCT 194

```

Db      14  GInGluAspGlyPheValAsnHisMetMetAsnAspGlyGluAspValGluAspLeuAsp 33
QY      195  TGAAGGCTGACACCGAGATTCTTGAGAGAGCAACCGGACAATATATACCGAGCTTTTGCAGC 254
Db      34  GlyGlyAlaGluGlnPheGluTyrAsp-GluAspGlyLysIleGlyLysArgLysAl 53
QY      255  AGCTATTTTGTAGAGCTTCTAGAGAAAGAGAGAGAAACCAACTTTGATCCAGCAATG 314
Db      53  aAlaLysLeuGlnAlaLysGluGluLysArgGlnMetArgGluTyrGluValArgGluAr 73
QY      315  GGGAGTAGAGTAGAAGACCGCTTCTATTAACAATCATGCTATCGAGGAGCAGAACACCACC 374
Db      73  gGluGluArgLysArgArgGluGluGluArgGluLysLysArgAspGluGluArgAlaLys 93
QY      375  TGAGAAAAGTGTATCTTAAACAAGAGAGTCTCAGATATCTGGGAAGGAGGAGAGACATC 434
Db      93  sGluGluAlaAspGluLysAlaGluGluGluGluArgLeuArgLysGluArgGluGlu---- 111
QY      435  AGTCACCATCTTAGACTCTTCGAGAGAGATAGCAAGAAAAGAGAGGTTGCTGCTCAA 494
Db      112  -----LysGluArgLysGluHisGluGluTyrLeuAlaMetLys 124
QY      495  AATCAAGCTGCTTCCTCGGGGACATAGCCAGAGAGGAGGCAAGAAAATGAAAACAAA 554
Db      124  sAlaSerPheAlaIle-----GluGluGluGly-----ThrAs 135
QY      555  TAGTCTTCAAAATGAGGAAAAGAGGAAAACAAGTAGGAGACACTGTTTTACCTCCAGGA 614
Db      135  pAlaIleGluGlyGluGlu--AlaGluAsnLeuIleArgAspPheValAspTyrVally 154
QY      615  ACATGAAAATAATCCAAATCCATCAACCTTCTTATTAATGTCATTCTTCTCTGAGGAA 674
Db      154  sThrAsnLysValValAsnIleAspGluLeuSer-----SerHisPheGlyLeuLysE 172
QY      675  GGAAGATTGTTGTTGTGAATAACATTCGTTACTGTTGTGA 717
Db      172  rGluAspAlaVal-----AsnArgLeuGlnHisPheIleGlu 184

RESULT 35
S55395
neurofilament protein M - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C:Accession: S55395
R:Vitadello, M.; Vettore, S.; Lamar, E.; Chien, K.R.; Gorza, L.
A:Description: Neurofilament mRNA and protein are expressed in precursors of heart condu
A:Reference number: S55395
A:Accession: S55395
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-644 <VT>
A:Cross-references: UNIPROT:P54938; EMBL:247378; NID:g854352; PIDN:CAA87454.1; PID:g8543
C:Superfamily: cytoskeletal keratin

Alignment Scores:
Pred. No.: 0.668 Length: 644
Score: 99.00 Matches: 48
Percent Similarity: 40.96% Conservative: 29
Best Local Similarity: 25.53% Mismatches: 70
Query Match: 6.03% Indels: 41
DB: 2 Gaps: 7

AF334735 (1-954) x S55395 (1-644)
QY      121  GTTCTTACCAAGAGATGTCGATTCCATTCCTCAACACCACCTACCGAATCCACAGGA 180
Db      172  LeuLeuAsnValLysMetAlaLeuAspIleGluIleAlaIleArg----- 187
QY      181  TTTGGGAATCTTCTTGAAGGG-----CTGACA 207
Db      188  -----LysLeuLeuGluGlyGluThrArgPheSerThrPheSerGlySerIleThr 205

```

```

QY      208  CGCAGATTCTGAGAGACCAACCGACATATATACAGCTTTTGACAGCAGCCTATTTTGAG 267
Db      206  GlyProLeuTyrThrHisArgGlnProSerVal----- 216
QY      268  AGCTTCTAGAGAAAGAGAGAAAACCAACTTTGTATCCAGCAGCAATGGGGAGTAAAGTA 327
Db      217  ThrIleSerSerLysIleGlnLysThrLysValGluAlaProLys-----LeuLysVal 234
QY      328  GAAGACCCCTTTCTATAACAATCATGCAATTCGAGGAGCAAGAACCACTCTGAGAAAAGTGT 387
Db      235  GlnHisLysPheValGluGluIleLeuGluThrLysValGluAspGluLysSerGlu 254
QY      388  CCTAAACAA-----GAACAGCTCAGATATCTGGGAAGGAGGAGCAG 429
Db      255  MetGluAspAlaLeuThrAlaIleAlaGluGluLeuAlaValSerValLysGluGluGlu 274
QY      430  ACATCAGTCACCATCTTAGACTCTTCTGAGGAAGATAGAAAAGAAAGAGAGGTTGCTGCT 489
Db      275  LysGluGluGluAlaGluGlyLysGluGluGlnGluAlaGluGluGluValAlaAla 294
QY      490  GTCAAAATCCAA-----GCTGCTTCGGGGACACATAGCCAGAGAGAGGCAAGAAA 543
Db      295  AlaLysLysSerProValLysAlaThrThrProGluIleLysGluGluGluGlyGluLys 314
QY      544  ATGAAAACAATAGTCTTCAAAATGAGGAAAAGGAAAACAAGTAGGAGCAGCTGTTT 603
Db      315  GluGlu-----GluGlyGlnGluGluGluGluGluGluGluGluGluGlyVa 330
QY      604  TACCTCCAGGAAACATGAAAA 625
Db      330  LysSerAspGlnAlaGluGlu 337

RESULT 36
A71623
probable secreted protein PFB0115w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C:Accession: A71623
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: A71623
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1192 <GAR>
A:Cross-references: UNIPROT:O96127; GB:AE001373; GB:AE001362; NID:g3845097; PIDN:AAC7181
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0115w

Alignment Scores:
Pred. No.: 0.721 Length: 1192
Score: 99.00 Matches: 38
Percent Similarity: 48.82% Conservative: 24
Best Local Similarity: 29.92% Mismatches: 42
Query Match: 6.03% Indels: 23
DB: 2 Gaps: 5

AF334735 (1-954) x A71623 (1-1192)
QY      277  GAGAAAAGAGAGAAAACCAACTTTGATCCAGCAATGGGGAGTAAAGTAGAACCGC 336
Db      416  GluAsnLysGluGluThrGluValAspGluLysLys---ThrGluLysAlaGluGlu--- 433
QY      337  TTCTATAACATCATGATTCGAGAGCAAGAACCACTCTGAGAAAAGTATCTCTAAACAA 396
Db      434  -----GluLeuGluGluAspLysGluGluSerGluLysAspLys 446
QY      397  GAAGAGTCTCAGATATCTGGGAAGGAGGAGACATCAGTCACCATCTTAGACTCTTCT 456

```



Db 447 GluGluSerGlu-----LysAspLysGluGluSerGluLysAspLysGluGluSer 463  
QY 457 GAGAAAGATAAGAAAGAAAGAGAGTTGCTGCTCAAAATCCAAAGTCCTCCGGGGA 516  
Db 464 GluLysAspLysGluLysThrGluGluAspGluGluLysThrGluLysGluThr 483  
QY 517 CACATAGCCAGAGAGAGGCA-----AAGAAATGAAA----- 549  
Db 484 GluValTyrLysLysGluThrAspValAspGluLysLysGluLysGluThrGlyGlu 503  
QY 550 ---ACAATAGTCTCAAAATGAGAAAGAGAGAAACAAGTGAGGACACTGTTTAC 606  
Db 504 GlyThrAspGluGluAspLysGluLysGluLysGluLysAsp-AspGluGluThrLysValG 523  
QY 607 CTCAGGAAACATGAAAA 625  
Db 523 uGluLysLysThrGluLys 529  
  
RESULT 37  
S45706  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 10-Dec-1994 #sequence\_revision 19-Apr-1996 #text\_change 24-Nov-1999  
C:Accession: S45706  
R;Peng, M.; Fan, H.; Kirtley, T.L.; Caswell, A.H.; Schwartz, A.  
FEBS Lett. 348, 17-20, 1994  
A:Title: Structural diversity of triadin in skeletal muscle and evidence of its existenc  
A:Reference number: S45704; MUID:94298946; PMID:8026576  
A:Accession: S45706  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-213 <PEN>  
C:Superfamily: histone H1  
  
Alignment Scores:  
Pred. No.: 0.648 Length: 213  
Score: 98.50 Matches: 48  
Percent Similarity: 44.39% Conservative: 35  
Best Local Similarity: 25.67% Mismatches: 71  
Query Match: 6.00% Indels: 33  
DB: 2 Gaps: 7  
  
AF334735 (1-954) x S45706 (1-213)  
QY 41 GAAAAACGGTTACCCAGCAACTAGAAAAAACACCGAACCGCGGCACAGCTCGGAGA 100  
Db 19 GluLysGlnAlaLysAlaLysIleGluLysGluLysGluLysValSerAla--AlaSerThrL 38  
QY 101 GAAAGGAGTTCCATAGGCAGTTCTTACCAGAGAGATGTCGATTCATTCACACACC 160  
Db 38 yLysAlaValProAlaLysLysGluLysThrLysThrValGluGluThrA 58  
QY 161 ACTACGAATTCACAGAGATTGGGAATCTTTGAAGGGCTGACACCGAGATTCTGA 220  
Db 58 rGlyGluLysProGlyLysIleSerSerValLysLys-----AspLysGluLysThrL 76  
QY 221 GAGAGAACCGGACATATACAGCTTTTGCACGACCTTTTGGAGCCCTTCTAGAGA 280  
Db 76 yGluLysGluValLysValProAla-----SerLeuLysGluL 89  
QY 281 AAGAGAGAAAAACCACTTTGATCCAGCAGATGGGGAGTAGAGTGGAGCCGCTTCT 340  
Db 89 yGlySerGluThrLysLysAsp-----GluLysThrSerLysProGluProGlnIleL 107  
QY 341 ATACAATCATGCTTCGAGGAGCAAGACCACTCTGAGAAAGTGTATCTTAAACAAGAG 400  
Db 107 yS-----LysGluLysProGlyLysGluLysValLysProLysProPioG 122  
QY 401 AGTCTCAGATATCTGGGAGGAGAA-----GAGACATCAGTCACCA 442  
Db 122 InProGlnIleLysLysGluLysGluLysProGluGlnAspIleMetLysProGluLysThrA 142  
QY 443 TCTTAGACTCTTCTGAGGAAGATAAGAAAAAGAGAGGTTGCTGTCTCAAAATCCAAG 502

Db 142 laLeuHisGlyLysProGluGluLysValLeuLysGlnValLysAlaValThrThrGluL 162  
QY 503 CTGCTTCCGGGACACATAGCCAGAGAGGAGCAAGAAATGAATAACATAGTCTTC 562  
Db 162 yS-----HisValLysProLysProAlaLysLys-----AlaG 173  
QY 563 AAAATGAGGAAAAAGAG 579  
Db 173 luHisGlnGluLysGlu 178  
  
RESULT 38  
T15827  
hypothetical protein C53C9.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Feb-2000  
C:Accession: T15827  
R;Bentley, D.  
submitted to the EMBL Data Library, June 1995  
A:Description: The sequence of C. elegans cosmid C53C9.  
A:Reference number: Z18413  
A:Accession: T15827  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-418 <BEN>  
A:Cross-references: EMBL:U28734; NID:g861255; PID:g861256; PIDN:AAAB52603.1; GSPDB:GN000  
A:Experimental source: strain Bristol N2; clone C53C9  
C:Genetics:  
A:Gene: CESP:C53C9.2  
A:Map position: X  
A:Introns: 25/3; 187/1; 310/2; 380/3; 405/3  
  
Alignment Scores:  
Pred. No.: 0.704 Length: 418  
Score: 98.50 Matches: 45  
Percent Similarity: 40.31% Conservative: 32  
Best Local Similarity: 23.56% Mismatches: 73  
Query Match: 6.00% Indels: 41  
DB: 2 Gaps: 7  
  
AF334735 (1-954) x T15827 (1-418)  
QY 76 GGAACCGCGCCACAGCTCGGAGAGAAAGGAGTTCCATAGCAGATTCTTACCAGAAG 135  
Db 244 GlySerGlyGlyPheLeuLysValArgAspValLeuProHisThrValGlyGlyLysAsp 263  
QY 136 ATGTCG-----ATTCATTC-----TCCAACACCCAC 162  
Db 264 ileGluGluGluLysSerGluGlyIleValProLeuGlnSerGlyThrAsn 283  
QY 163 TACCGAATTCACAA-----GGATTGGG-----AATCTTCTT 195  
Db 284 LysLeuAlaSerGlnArgGlyMetThrGlyPheGlyThrProArgAsnThrGlnLeuArg 303  
QY 196 GAAGGGCTGACACGCGAGATTCTGAGAGCAACCGGACATATACCAGCTTTTCGACGA 255  
Db 304 AlaGlyTyrLysGluTrpIleGluAsp----- 313  
QY 256 GCCTATTTGAGAGCTTCTAGAGAAAGAGAGAAACCAACTTGTATCCAGCAGANTGG 315  
Db 314 -----TyrGluAlaLeuLysGluTrpGluGluThrLys-----ProGlyLys 329  
QY 316 GGGAGTAGGTAGAAGCCGCTTCTATAACATCATGCATTCGAGGAGCAAGACCACT 375  
Db 330 AlaSerSerValAspProPheGlyHisTyrLysLysLysPheGluGluArgGlu----- 347  
QY 376 GAGAAAGTGCCTCTAAACAAGAGAGTCTCAGATATCTGGGAAGAGGAGAGAGACATCA 435  
Db 348 -----SerSerArgGlnSerGluIleAspSerGlnSerValLysAlaSer 362  
QY 436 GTCACATCTTAGACTCTTCTCGAGAAAGATAAGAAAAAGAGAGGTTCTGCTGTCACAA 495  
Db 363 GluProValGluProGluGluGluGluGluGluGluGluGluGluGluGluGlu 382



```

QY 368 AACACCTGAGAAAAGTGTATCTTAAACAAGAGAGCTCTCAG-----A 409
Db 760 luthrlysgluGluGlyLysProLeuGlnGlnGluLysGluLysGluLysGlyG 780
QY 410 TATCTGGAGGAGGAGAGACATCATCTAGCATCTTCTGAGGAGATAGG 469
Db 780 luthrlysgluGluGlyLysProLeuGlnGlnGluLysGluLysGlyGlySera 796
QY 470 AAAAAGAGAGGTTGCTGCT-----GTCAAAATCCAAG 502
Db 796 rglYsgluAepIleAlaValAsnGlyGluValGluGlyLysGluValGluGlnGluT 816
QY 503 CTCCTCTCCGGGACACATAGCCAGAGAGAGCAAGAAATGAAACAAATAGTCTTC 562
Db 816 hrlysgluLysGlySer--GlyArgGluGluGlyLysGlyValValThrAsnGlyLeuA 835
QY 563 AAAATGAGGAAAAGAGGAAAACAACAGTGAAG 594
Db 835 splEuserProAlaAepGluLysGlyGly 845

RESULT 41
S60254
nuclear receptor co-repressor N-CoR - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 09-Jul-2004
C:Accession: S60254
R:Hoerlein, A.J.; Naeae, A.M.; Heinzel, T.; Torchia, J.; Gloss, B.; Kurokawa, R.; Ryan,
A;Title: Ligand-independent repression by the thyroid hormone receptor mediated by a nu
A;Reference number: S60254; MUID:96008539; PMID:7566114
A;Accession: S60254
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-2453 <HOB>
A;Cross-references: UNIPROT:Q60974; EMBL:U35312; NID:g1022717; PIDN:AAB17125.1; PID:g102

Alignment Scores:
Pred. No.: 0.977 Length: 2453
Score: 98.00 Matches: 45
Percent Similarity: 39.34% Conservative: 27
Best Local Similarity: 24.59% Mismatches: 63
Query Match: 5.97% Indels: 48
DB: 2 Gaps: 9

AF334735 (1-954) x S60254 (1-2453)
QY 67 AAAACAACCGACCGCGCCAGCTCGAGAGAAAGAGGTTCCATAGGAGT--- 123
Db 457 LysAsnPhedGlyLeuIleAlaSerTyrLeuGluArgLysSerValProAspCysValLeu 476
QY 124 -----CTTACCAGAGATGTCGATTCCATTCTCAACACCCACTACCGA----- 168
Db 477 TyrTyrTyrLeuThrLysLys-----AsnGluAsnTyrLysAlaLeu 490
QY 169 ATTCCAAGAGATTTCGGAATCTTCTTGAAGGCTCAGCGAGATTCTGAGAGACAA 228
Db 491 ValArgAsnTyrGlyLys--ArgArgGlyArgAsnGlnGlnIleAlaArgProSer 509
QY 229 CCGACACATATACAGCTTTTGACGAGCCTATTTCGAGAGCCTCTTACGAGAAAGAGAG 288
Db 510 GlnGlu-----GluLysValGlu 515
QY 289 AAAACCAACTTTTCATCCAGCAGAAATGGGAGTAAGTAGAAGACCGCTTCTATAACAAT 348
Db 516 GluLysGluLysAspLysAlaGluLysThrGluLysLysGlu----- 529
QY 349 CATGCAATTCGAGGAGCAACACCTCTGAGAAAGTATGCTTAAACAAGAGAGTCTCAG 408
Db 530 -----GluGluLysLysAspAspGluGluLysAspAspLysGluAspSerLysGlu 546
QY 409 ATATCTGGGAGGAGAGAGACATCAGTCACCATCTTACACTCTTCTGAGGAGATAG 468
Db 547 ThrThrLysGluLysAspArgThrGluAlaThr-----AlaGluGluProGlu 562

```

```

QY 469 GAAAAGAGAGAGTGTCT-----GCTGTCAAATCCAGCTCGCTTCCGG 513
Db 563 GluArgGluGlnValThrProArgGlyArgLysThrAlaAsnSerGlnGlyArgGlyLys 582
QY 514 GGACACATAGCCAGA-----GAGGAGGCAAGAAATGAAACAAATAGTCTTCAA 564
Db 583 GlyArgValThrArgSerMetThrSerGluAlaAlaAlaAlaAsnAlaAlaAlaAla 602
QY 565 AATGAGGAA 573
Db 603 ThrGluGlu 605

RESULT 42
S49313
protein kinase - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C>Date: 16-Feb-1995 #sequence_revision 12-May-1995 #text_change 16-Aug-2004
C:Accession: S52076; S49313
R:Wetterauer, B.W.; Hamker, U.; von Haeseler, A.; MacWilliams, H.K.; Simon, M.N.; Veron,
Biochim. Biophys. Acta 1265, 97-101, 1995
A;Title: A protein kinase from Dictyostelium discoideum with an unusual acidic repeat d
A;Reference number: S52076; MUID:95161460; PMID:7857991
A;Accession: S52076
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1094 <WR2>
A;Cross-references: UNIPROT:Q23915; EMBL:Z37981; NID:g551445; PIDN:CRA86053.1; PID:g551
C:Genetics:
A;Introns: 35/3; 104/1; 166/2
C;Superfamily: protein kinase homology
C;Keywords: ATP
F:20-281/Domain: protein kinase homology <KIN>
F:28-36/Region: protein kinase ATP-binding motif

Alignment Scores:
Pred. No.: 0.984 Length: 1094
Score: 97.50 Matches: 44
Percent Similarity: 41.38% Conservative: 40
Best Local Similarity: 21.67% Mismatches: 64
Query Match: 5.94% Indels: 55
DB: 2 Gaps: 8

AF334735 (1-954) x S49313 (1-1094)
QY 38 AAGAAAACCGTTTACCAGCACTAGAAAACACCGACCGCGCCAGCTCG 97
Db 558 GluGluGluGluTyrGlnArgGlnLeuGlnGlnGlnGlnGlnGlnGlnGln 577
QY 98 AGAGAAAGGAGGTTCCATAGGAGGTTCTTAC----- 128
Db 578 GlnGlnGlnHisGlnHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 597
QY 129 -CAAGAGATGTCGATTCTTCCATCTCCACACCCTACCGAATTTCCACAGGATTGGGA 187
Db 597 uGluGluGluTyrAspAspValIleArgHis----- 607
QY 188 ATCTTCTTGAAGGCTGACCGGAGATTCTCAGAGAGCAACCGGACCAATATACAGCTT 247
Db 608 -----AspThrAspSerGluGlu-GluSerLysAspLysThrProLeuP 622
QY 248 TTGACGAGCCTTATTTGAGAGCCTTTAGAGAAAAGAGAGAAAACCAACTTTTGATCCAG 307
Db 622 rotPAspGlnHisPheGlu-----LysGlnLysGluSerGluAsn----- 635
QY 308 CAGAATGGGGAGTAGTAGTAGAAGACCGCTTCTATAAC-----AATCATGCAT 355
Db 636 -----LysValGluGlnGluGluThrAsnValValValAlaAlaAsnSerGlnG 651
QY 356 TCGAGGAGCAAGAACCCCTGAGAAAAGTGTCTTAAACAAGAGAGTCTCAGATATCTG 415
Db 651 luthrAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 671

```





R:Johnson, D.; Wamsley, P.; Bradshaw, H.  
submitted to the EMBL Data Library, February 1997  
A:Description: The sequence of *C. elegans* cosmid ZK354.  
A:Reference number: Z20120

A:Accession: T25991

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-224 <JOH>

A:Cross-references: UNIPROT:P91567; EMBL:U88172; PIDN:AB42259.1; GSPDB:GN00022; CESP:ZK

A:Experimental source: strain Bristol N2; clone ZK354

C:Genetics:

A:Gene: CESP:ZK354.7

A:Map position: 4

A:Introns: 25/3

Alignment Scores:

Pred. No.: 1 Length: 224  
Score: 96.50 Matches: 35  
Percent Similarity: 45.24% Conservatives: 22  
Best Local Similarity: 27.78% Mismatches: 56  
Query Match: 5.88% Indels: 13  
DB: 2 Gaps: 4

AF334735 (1-954) x T25991 (1-224)

QY 229 CCGGCAATATACAGCTTTTCAGCAGCCTATTTTGAGAGCCTTCTAGAGAAAGAGAG 288

DB 92 ProGluAsnAsnLysGlnTrpPheAlaPheTyTr-----Ile 104

QY 289 AAAACCACTTTGAT---CCAGCAGAA-----TGGGGAGTAGAGTAGAAGAC----- 333

DB 105 LysThrLysGluAspGlyProAlaArgAspAlaTrpLysAsnHisLysProAspGlyMet 124

QY 334 ---CGCTTCTATACAATCATCCTCGAGGAGCAAGAACCACTGAGAAAGTAGTCCT 390

DB 125 LysArgIleTyIleSerPheLysGlyAspGluLysGlyAspGluLysAspGlu 144

QY 391 AAACAGAGAGCTCTCAGATATCTGGAGAGGAGAGACATCAGTCACCATCTTAGAC 450

DB 145 LysLysGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 164

QY 451 TCTTCTGAGGAGATAGAGAAAGAGAGGTTCGTCTGCTCAAAATCCAAAGCTCCCTC 510

DB 165 LysLysGluAlaLysLysGluLysLysGluLysLysGluLysLysGluLys 184

QY 511 CGGGACACATGACGAGAGAGCAAGAAATAAATAAGTAGTCTTCAAAATGAG 570

DB 185 LysLysGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 204

QY 571 GAAAAGAGAGAAACAAG 588

DB 205 LysLysGluLysLys 210

RESULT 48

A45605

A:Title: Repeat structures in a Plasmodium falciparum parasite (Plasmodium

C:Species: Plasmodium falciparum

C:Date: 22-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004

C:Accession: A45605; A54517

R:Coppel, R.L.

Mol. Biochem. Parasitol. 50, 335-347, 1992

A:Title: Repeat structures in a Plasmodium falciparum protein (MESA) that binds human ex

A:Reference number: A45605; MUID:92158014; PMID:1741020

A:Accession: A45605

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1526 <COP>

A:Experimental source: Papua New Guinea isolate FC27

A:Note: sequence extracted from NCBI backbone (NCBI:83648, NCBIP:83656)

R:Coppel, R.L.; Culvenor, J.G.; Bianco, A.E.; Cretzner, P.E.; Stahl, H.D.; Brown, G.V.;

Mol. Biochem. Parasitol. 20, 265-277, 1986

A:Title: Variable antigen associated with the surface of erythrocytes infected with matu

A:Reference number: A54517; MUID:87014571; PMID:3531849

A:Accession: A54517

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 222-443, 'K', 445 <CO2>

A:Cross-references: GB:M15319; NID:g160060; PID:g552170

C:Superfamily: ring-infected erythrocyte surface antigen; dnaJ amino-terminal homology

C:Keywords: surface antigen; tandem repeat

Alignment Scores:

Pred. No.: 1.27 Length: 1526  
Score: 96.50 Matches: 46  
Percent Similarity: 41.15% Conservatives: 33  
Best Local Similarity: 23.96% Mismatches: 57  
Query Match: 5.88% Indels: 56  
DB: 2 Gaps: 8

AF334735 (1-954) x A45605 (1-1526)

QY 42 AAAAAACGGTTTACCCAGCACTAGAAAAACAACCGGCGCGCACCAGCTCGGAGAG 101

DB 874 LysLysArgValLysLys-----ArgAsnAsnLysAsnGluArg----- 886

QY 102 AAGGAGGTTTCCATAGGCAGTTCT---TACCAAGAAGATGTCGATTTCATTTCCAAACAC 158

DB 887 LysAspAsnValIleGlyLysGluLysMetLysGluAspValAsnGlu----- 902

QY 159 CCCTACCGAATCCACAGGATTTGGGAATCTTCTTGAAGGGCTGACACCGGAGATTCT 218

DB 903 -----LysAspThrAlaAsnLysAsp-LysGluIleG 913

QY 219 GAGAGACCAACCGCAATATATACAGCTTTTTCGAGCAGCCTATTTTTCGAGACCTCTCTAGA 278

DB 913 uGlnGluLysGluLys-----GluGluValLysG 923

QY 279 GAAAGAGAGAAAAACCAACTTTTGATCCAGCAGAAATGGGGAGTAGAGTAGAACCGCTT 338

DB 923 uLysGluValLysGluLysGluValLysGluLysGluLysGluLysGluLysGluLys 942

QY 339 CTATAACATCATGCTTCGAGGAGCAAGAACCACTGAGAAAAGTAGTCTCTAAACAAGA 398

DB 943 -----GluGluValLysGluLysGluValLysGluLysGluLysGluLysGluLys 960

QY 399 AGAGTCTCAGATATCTGGAGAGGAGAGACATCAGTCACCATCTTAGACTCTCTGA 458

DB 960 AspThrGluSerLysAspLysGlu-----IleG 970

QY 459 GGAAGATAAGAAAAAGAGAGGTTGCTGTGCTCAAAATC----- 498

DB 970 uGlnGluLysGluLysGluValLysGluValLysGluLysAspThrGluAsnLysAs 990

QY 499 -----CAAGCTGCTTCGCGGGACACATAGCCAGAGAGGAGGCAAGAAAT 545

DB 990 pLysValIleGlyGlnGluIleIleGluGluLysLysGluValLysLysArgVa 1010

QY 546 GAAACAAATAGTCTTCAAAATAGAGAAAGAG 579

DB 1010 lLysLysArgAsnAsnLysAsnGluAsnLysAsp 1021

RESULT 49

B64171

A:Title: Repeat structures in a Plasmodium falciparum protein (MESA) that binds human ex

C:Species: Haemophilus influenzae

C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 16-Aug-2004

C:Accession: B64171

R:Flaischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, J.

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.

; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

Science 269, 456-512, 1995

A:Authors: Ghem, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: B64171







```
Db      |||||:::
511 AlaLysGluGluGluGlyGlyGluGlyGluGlnGlyGluGluThrLysGluAlaGluGlu 530
QY      |||||:::
568 GAGGAAAAAGAGGAA 582
Db      |||||:::
531 GluGluLysLysAsp 535
```

Search completed: February 15, 2005, 10:07:41  
Job time : 61.5 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 15, 2005, 09:49:49 ; Search time 161.5 Seconds  
(without alignments)  
6049.827 Million cell updates/sec

Title: AF334735  
Perfect score: 1641  
Sequence: 1 TCGCCCTTCTCGCGCGCG.....AAAAAAAAAAAAAAAAAAAA 954

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO spool\_p/AF334735/runat\_15022005\_094942\_10092/app\_query.fasta.1.1095  
-DB=UniProt\_03 -QFWT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=50 -MODE=LOCAL  
-OUTFWT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=AF334735 @CGN 1.1 244 @runat\_15022005\_094942\_10092 -MCFU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -D5PBLOCK=100 -LONGLOG  
-DSV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 03:1  
1: uniprot\_sprot:1  
2: uniprot\_trembl:1

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	812	49.5	163	1 SP17 PAPA	Q95230 papio hamad
2	778	47.4	151	1 SP17 HUMAN	Q15506 homo sapien
3	758	46.2	151	1 SP17 MACPA	Q19021 macaca fasc
4	710	43.3	143	2 Q9BDQ6	Q9bdq6 macaca mula
5	684.5	41.7	212	2 Q9XT29	Q9xt29 callithrix
6	589	35.9	147	2 Q9TU08	Q9tu08 ovula aries
7	569	34.7	146	1 SP17 RABIT	P36425 oryctolagus
8	561.5	34.2	148	2 Q9ZIK2	Q9zik2 rattus norv
9	558	34.0	149	1 SP17 MOUSE	Q62252 mus musculus
10	544	33.2	153	1 SP17 MACEU	Q62770 macropus eu
11	502.5	30.6	179	1 SP17 MONDO	Q62771 monodelphis
12	491.5	30.0	141	2 Q6SJS2	Q6sj92 ornithorhyn
13	364	22.2	495	2 Q6DIJ9	Q6di99 xenopus tro
14	134	8.2	189	2 Q20199	Q20199 caenorhabdi
15	131	8.0	422	2 Q6WN93	Q6wn93 brachydanio
16	122.5	7.5	368	2 Q25114	Q25114 hemicentrot

17	122.5	7.5	369	1	KAPR_STRPU	Q26619 strongyloce
18	121.5	7.4	221	2	Q9HAY3	Q9hay3 homo sapien
19	121.5	7.4	379	2	Q9HAY5	Q9hay5 homo sapien
20	121.5	7.4	493	2	Q75952	Q75952 homo sapien
21	121	7.4	596	2	Q9CM47	Q9cm47 macaca fasc
22	119.5	7.3	562	2	Q8BU41	Q8bu41 debaryomyce
23	118.5	7.2	382	2	Q9BUB1	Q9bub1 homo sapien
24	118.5	7.2	403	1	KAP2_HUMAN	P13861 homo sapien
25	116.5	7.1	216	2	Q7S160	Q7s160 neurospora
26	116.5	7.1	401	2	Q8K1M2	Q8k1m2 rattus norv
27	116.5	7.1	498	2	Q710D7	Q710d7 vulpes vulp
28	115.5	7.0	400	1	KAP2_BOVIN	P00515 bos taurus
29	115	7.0	532	2	Q9T069	Q9t069 arabidopsis
30	114	6.9	992	2	Q7S5R9	Q7s5r9 neurospora
31	113.5	6.9	443	2	Q9LYR0	Q9lyr0 arabidopsis
32	112.5	6.9	3901	2	Q9N533	Q9n533 caenorhabdi
33	112	6.8	845	2	Q8NE71	Q8ne71 homo sapien
34	112	6.8	3290	2	Q7RRX0	Q7rrx0 plasmodium
35	111.5	6.8	396	2	Q6TU32	Q6tu32 aplysia cal
36	111.5	6.8	446	2	Q9LVA0	Q9lva0 arabidopsis
37	111	6.8	402	2	Q8K1M3	Q8k1m3 mus musculu
38	110.5	6.7	134	2	Q9CYV3	Q9cyv3 m mus muscu
39	110.5	6.7	921	2	Q8QTD2	Q8qtd2 white spot
40	110.5	6.7	922	2	Q9VAV5	Q9vav5 white spot
41	110.5	6.7	922	2	Q9ILB5	Q9ilb5 white spot
42	110	6.7	103	2	Q9SD11	Q9sd11 arabidopsis
43	109	6.6	312	2	Q6CCN1	Q6ccn1 yarrowia li
44	108	6.6	403	1	KAPR_BLAEM	P31320 blastoclad
45	108	6.6	1132	1	YKK5_YEAST	P34250 saccharomyc

#### ALIGNMENTS

##### RESULT 1

SP17\_PAPHA  
ID SP17\_PAPHA STANDARD; PRT; 163 AA.  
AC Q95230;  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Sperm surface protein SP17 (Sperm autoantigenic protein 17).  
GN Name=SP17; Synonyms=SP17;  
OS Papio hamadryas (Hamadryas baboon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;  
OC Cercopitheciinae; Papio.  
OX NCBI\_TaxID=9557;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=97264490; PubMed=9110316;  
RX DOI=10.1002/(SICI)1098-2795(199705)47:1-66::AID-MRD9>3.0.CO;2-O;  
RA Adoyo P.A., Lea I.A., Richardson R.T., Widgeon E.E., O'Rand M.G.;  
RT "Sequence and characterization of the sperm protein SP17 from the baboon."  
RL Mol. Reprod. Dev. 47:66-71(1997).  
CC -!- FUNCTION: Sperm surface zona pellucida binding protein. Helps to bind spermatozoa to the zona pellucida with high affinity. Might function in binding zona pellucida and carbohydrates (By similarity).  
CC -!- SUBUNIT: Homodimer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Membrane-associated (Potential).  
CC -!- TISSUE SPECIFICITY: Testis- and sperm-specific.  
CC -!- SIMILARITY: Contains 1 IQ domain.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

```

DR EMBL; U75209; AAB38534.1; -.
DR HSSP; P12367; 1R2A.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR003117; RIIA.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF02197; RIIa; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00394; RIIa; 1.
DR PROSITE; PS50096; IQ; 1.
KW Membrane.
FT DOMAIN 114 143 IQ.
SQ SEQUENCE 163 AA; 18755 MW; A890B59B3F6469BD CRC64;

Alignment Scores:
Pred. No.: 4,35e-61 Length: 163
Score: 812.00 Matches: 158
Percent Similarity: 98.1% Conservative: 3
Best Local Similarity: 96.34% Mismatches: 2
Query Match: 49.48% Indels: 1
DB: 1 Gaps: 0

AF334735 (1-954) x SP17_PAPHA (1-163)
QY 136 ARTGCGATTCCATTCTCCACACCCACTACCGAATTCACAGGATTGGGAATCTTCTT 195
DB 1 MetSerIleProPheSerAsnThrHisTyArgIleProGlnGlyPheGlyAsnLeuLeu 20
QY 196 GAAGGCTGACAGCGGAGATTCTGAGAGAGCAACCGGACAATATACCAAGCTTTTCAGCA 255
DB 21 GluGlyLeuThrArgGluLeuArgGluGlnProAspAsnIleProAlaPheAlaIa 40
QY 256 GCCTATTTGAGAGCTTCTAGAGAAAGAGAGAAACCACTTTGATCCAGCAGAAATGG 315
DB 41 AlatyPheGluSerLeuLeuGlnysArgGluLysThrAsnPhaAspProAlaGluTrp 60
QY 316 GGGAGTAGGTAGAGACCGCTTCTATACCAATCATGCTTCGAGGAGCAAGAACCACT 375
DB 61 GlySerLysValGluAspArgPheTyAsnAsnHisAlaPheGluGlnGluProPro 80
QY 376 GAGAAAGTGATCTTAACAGAGAGCTTCAGATATCTGGAGAGAGAGAGACATCA 435
DB 81 GluLysSerAspProLysGlnGluSerGlnValSerGlyLysGluGluThrSer 100
QY 436 GTCCACTTGTAGCTTCTTCAGAGAGATAGGAAAGAGAGAGAGGTTGCTCTCTCAA 495
DB 101 ValThrIleLeuAspSerSerGluGluAspLysGluLysGluGluValAlaAlaVallys 120
QY 496 ATCCAGCTGCCTTCGGGGGACATAGCCAGAGAGAGGAGGCAAGAAATGAAACAAAT 555
DB 121 IleGlnAlaAlaPheArgGlyHisValAlaArgGluGluValLysMetLysThrAsp 140
QY 556 ACTGTTCAAAATGAGAAAGAGAGAAACAAAGTGAAGACACTGTTTACTCCAGGAA 615
DB 141 SerLeuGlnAsnGluGluLysGluGluAsn-SerGluAspThrGlyPheThrSerArgTh 160
QY 616 ACATCAAAA 625
DB 160 rHisGluLys 163

RESULT 2
SP17_HUMAN
ID SP17_HUMAN STANDARD; PRT; 151 AA.
AC Q15506; Q9EXF7;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Sperm surface protein Sp17 (Sperm autoantigenic protein 17) (Sperm
GN Name=SP17; Synonyms=SP17;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=96305346; PubMed=8688458; DOI=10.1016/0167-4781(96)00077-2;
RA Lea I.A., Richardson R.T., Widgren E.E., O'Rand M.G.;
RT "Cloning and sequencing of cDNAs encoding the human sperm protein,
RL Sp17.";
RN [2]
RP Biochim. Biophys. Acta 1307:263-266(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=22280349; PubMed=12393185; DOI=10.1016/S0167-4781(02)00478-5;
RA Buchli R., De Jong A., Robbins D.L.;
RT "Genomic organization of an intron-containing sperm protein 17 gene
RL model.";
RN [4]
RP Biochim. Biophys. Acta 1578:29-42(2002).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Myers R.M.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Skalska U., Smalls D.E.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RN and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC FUNCTION: Sperm surface zona pellucida binding protein. Helps to
CC bind spermatozoa to the zona pellucida with high affinity. Might
CC function in binding zona pellucida and carbohydrates (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Membrane-associated (Potential).
CC -!- TISSUE SPECIFICITY: Testis- and sperm-specific.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z48570; CAA88459.1; -.
CC EMBL; AF334735; AAK20878.1; -.
CC EMBL; AF334810; AAK28125.1; -.
CC EMBL; BC032457; AAK32457.1; -.
CC FIR; J38243; J38243.
CC HSSP; P12367; 1R2A.
CC Genew; HGNC:11210; SPAL17.
CC H-InvdB; HIX0010221; -.
CC MIM; 608621; -.
CC GO; GO:0007338; P:fertilization (sensu Animalia); TAS.
CC GO; GO:0007283; P:spermatogenesis; TAS.
CC InterPro; IPR000048; IQ region.
CC InterPro; IPR003117; RIIa.
CC Pfam; PF00612; IQ; 1.
CC Pfam; PF02197; RIIa; 1.

```

```

DR SMART; SM00015; IQ; 1.
DR SMART; SM00394; RIIA; 1.
DR PROSITE; PS50096; IQ; 1.
KW Membrane.
FT DOMAIN 114 143 IQ.
SQ SEQUENCE 151 AA; 17406 MW; BC99EC310FA0E54A CRC64;

Alignment Scores:
Pred. No.: 3,62e-58 Length: 151
Score: 778.00 Matches: 151
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 47.41% Indels: 0
DB: 1 Gaps: 0

AF334735 (1-954) x SPI7_HUMAN (1-151)

QY 136 ATGTCGATTCATTCCTCAACACCCACTACCGAATTCACAGATTGGGAATCTTCTT 195
Db 1 MetSerIleProPheSerAsnThrHisTyrArgIleProGlnGlyPheGlyAsnLeu 20

QY 196 GAAGGGCTGACCGCAGATTCTGAGAGCAACCGACAATATACAGCTTTTGCGACA 255
Db 21 GluGlyLeuThrArgGluIleLeuArgGluGlnProAspAsnIleProAlaPheAla 40

QY 256 GCCTATTGAGAGCCTTCTAGAGAAAGAGAGAAACCAACTTTGATCCAGCAGATCG 315
Db 41 AlaTyrPheGluSerLeuLeuGluLysArgGluLysThrAsnPheAspProAlaGluTrp 60

QY 316 GCGAGTAAGTAAAGACCGCTTCTAGAGAAAGAGAGAAACCAACTTTGATCCAGCAGATCG 375
Db 61 GlySerIleValGluAspArgPheTyrAsnAsnHisAlaPheGluGluGlnProPro 80

QY 376 GAGAAAGTGAATCTTAAACAAGAGAGTCTCAGATATCTGGGAAGGAGAGAGATCA 435
Db 81 GluLysSerAspProLysGlnGluSerGlnIleProGlyLysGluGluAlaSer 100

QY 436 GTCACCATCTTAGACTCTTCTGAGCAAGTAAAGAGAGAGAGAGAGAGAGAGAGAG 495
Db 101 ValThrIleLeuAspSerSerGluGluAspLysGluLysGluValAlaAlaValLys 120

QY 496 ATCCAAAGCTGCTTCCGGGGACACATAGCCAGAGAGAGAGAGAGAGAGAGAGAG 555
Db 121 IleGlnAlaAlaPheArgGlyHisValAlaAlaArgGluGluValLysMetLysThrAsp 140

QY 556 AGTCTTCAAAATCAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 588
Db 141 SerLeuGlnAsnGluGluLysGluLysGluLysGluLysGluLysGluLysGluLys 151

RESULT 3
SPI7_MACFA
ID SPI7_MACFA STANDARD; PRT; 151 AA.
AC Q19021;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Sperm surface protein Spi7 (Sperm autoantigenic protein 17).
GN Name=SPI7; Synonyms=SPI7;
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OC NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Lea I.A., O'Rand M.G.;
RT "The immune response to immunization with sperm antigens in the
RT Macaque oviduct."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Sperm surface zona pellucida binding protein. Helps to
CC bind spermatozoa to the zona pellucida with high affinity. Might
CC function in binding zona pellucida and carbohydrates (By

```

```

similarity).
-!- SUBUNIT: Homodimer (By similarity).
-!- SUBCELLULAR LOCATION: Membrane-associated (Potential).
-!- TISSUE SPECIFICITY: Testis- and sperm-specific.
-!- SIMILARITY: Contains 1 IQ domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AF005551; AAB62888.1; -.
DR HSP; P12367; IR2A.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR003117; RIIA.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF02197; RIIA; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00394; RIIA; 1.
DR PROSITE; PS50096; IQ; 1.
KW Membrane.
FT DOMAIN 114 143 IQ.
SQ SEQUENCE 151 AA; 17401 MW; E681586AB713F03C CRC64;

Alignment Scores:
Pred. No.: 1,89e-56 Length: 151
Score: 758.00 Matches: 146
Percent Similarity: 98.01% Conservative: 2
Best Local Similarity: 96.69% Mismatches: 3
Query Match: 46.19% Indels: 0
DB: 1 Gaps: 0

AF334735 (1-954) x SPI7_MACFA (1-151)

QY 136 ATGTCGATTCATTCCTCAACACCCACTACCGAATTCACAGATTGGGAATCTTCTT 195
Db 1 MetSerIleProPheSerAsnThrHisTyrArgIleProGlnGlyPheGlyAsnLeu 20

QY 196 GAAGGGCTGACCGCAGATTCTGAGAGCAACCGACAATATATACAGCTTTTGCGACA 255
Db 21 GluGlyLeuThrArgGluIleLeuArgGluGlnProAspAsnIleProAlaPheAla 40

QY 256 GCCTATTGAGAGCCTTCTAGAGAAAGAGAGAAACCAACTTTGATCCAGCAGATCG 315
Db 41 AlaTyrPheGluSerLeuLeuGluLysArgGluLysThrAsnPheAspProAlaGluTrp 60

QY 316 GCGAGTAAGTAAAGACCGCTTCTAGAGAAAGAGAGAAACCAACTTTGATCCAGCAGATCG 375
Db 61 GlySerIleValGluAspArgPheTyrAsnAsnHisAlaPheGluGluGlnProPro 80

QY 376 GAGAAAGTGAATCTTAAACAAGAGAGTCTCAGATATCTGGGAAGGAGAGAGATCA 435
Db 81 GluLysSerAspProLysGlnGluSerGlnIleProGlyLysGluGluAlaSer 100

QY 436 GTCACCATCTTAGACTCTTCTGAGCAAGTAAAGAGAGAGAGAGAGAGAGAGAGAG 495
Db 101 ValThrIleLeuAspSerSerGluGluAspLysGluLysGluValAlaAlaValLys 120

QY 496 ATCCAAAGCTGCTTCCGGGGACACATAGCCAGAGAGAGAGAGAGAGAGAGAGAG 555
Db 121 IleGlnAlaAlaPheArgGlyHisValAlaAlaArgGluGluValLysMetLysThrAsp 140

QY 556 AGTCTTCAAAATCAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 588
Db 141 SerLeuGlnAsnGluGluLysGluLysGluLysGluLysGluLysGluLysGluLys 151

RESULT 4
Q9BDQ6
ID Q9BDQ6 PRELIMINARY; PRT; 143 AA.
AC Q9BDQ6;

```

DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Sperm protein 17 (Fragment).  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Macaca.  
 OX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Splice;  
 RX MEDLINE=22280349; PubMed=12393185; DOI=10.1016/S0167-4781(02)00478-5;  
 RA Buchli R., De Jong A., Robbins D.L.;  
 RT "Genomic organization of an intron-containing sperm protein 17 gene (Sp17-1) and an intronless pseudogene (Sp17-2) in humans: a new model.";  
 RT Biochim. Biophys. Acta 1578:29-42 (2002).  
 RL EMBL; AF334809; AAK28124.1; -;  
 DR HSP; P12367; 1R2A.  
 DR GO; GO:0008603; F:CAMP-dependent protein kinase regulator act. . . ; IEA.  
 DR GO; GO:0007165; P:signal transduction; IEA.  
 DR InterPro; IPR000048; IQ region.  
 DR InterPro; IPR003117; RIIA.  
 DR Pfam; PF00612; IQ; 1.  
 DR Pfam; PF02197; RIIA; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00394; RIIA; 1.  
 DR PROSITE; PS50096; IQ; 1.  
 FT NON TER 1 1  
 FT NON TER 143 143  
 SQ SEQUENCE 143 AA; 16424 MW; 70B538F7F876B465 CRC64;

Alignment Scores:  
 Pred. No.: 2,52e-52 Length: 143  
 Score: 710.00 Matches: 137  
 Percent Similarity: 97.20% Conservative: 2  
 Best Local Similarity: 95.80% Mismatches: 4  
 Query Match: 43.2% Indels: 0  
 DB: 2 Gaps: 0

AF334735 (1-954) x Q9BDQ6 (1-143)  
 QY 154 RACCCCTACCGATCCCAAGGATTTGGATCTTCTTGAAGGCTGACAGCGAG 213  
 DB 1 AsnThrHisTyrArgIleProGlnGlyPheGlyAsnLeuLeuGluGlyLeuThrArgGlu 20  
 QY 214 ATTCTGAGAGACCAACCGACATATACCGCTTTTTCAGCAGCGCTATTTGAGAGCCTT 273  
 DB 21 IleLeuArgGluGlnProAspAsnIleProAlaPheAlaAlaAlaTyrPheGluSerLeu 40  
 QY 274 CTAGAGAAAGAGAGAAACCACTTTGATCCAGAGATGGGGAGTAAGGTAGAGAC 333  
 DB 41 LeuGluLysArgGluLysThrAsnPheAspProAlaGluTrpGlySerLysValGluAsp 60  
 QY 334 CGCTCTTAACAATCATGCTTCGAGGAGCAAGACCTGAGAGAAAGTATCTTAA 393  
 DB 61 ArgPheTyrAsnAsnHisAlaPheGluGluGlnGlyProProGluLysSerAspProLys 80  
 QY 394 CAAGAAGAGCTCTCAGATATCTCGGAGGAGGAGAGACATCAGTCACCATCTTAGACTCT 453  
 DB 81 GlnGluGluSerGlnIleProGlyLysGluGluGluAlaSerValThrIleLeuAspSer 100  
 QY 454 TCTGAGGAAGATAGAGAAAGAGAGGTTGCTGTGTCGTCGTCGTCGTCGTCGTCG 513  
 DB 101 SerGluGluAspLysGluGluValAlaAlaValLysIleGlnAlaAlaPheArg 120  
 QY 514 GCACATAGCCAGAGAGAGCAAGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 573  
 DB 121 GlyHisValAlaArgGluGluValLysLysMetLysThrAspSerLeuGlnAsnGluGlu 140  
 QY 574 AAAGAGGAA 582

Db 141 LysGluGlu 143  
 RESULT 5  
 Q9XT29 PRELIMINARY; PRT; 212 AA.  
 ID Q9XT29  
 AC Q9XT29  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Sperm protein Sp17.  
 OS Callithrix jacchus (Common marmoset).  
 OC Eukaryota; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.  
 OX NCBI\_TaxID=9483;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RL Lea I.A., O'Rand M.G.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF134585; AAD30113.1; -;  
 DR HSP; P12367; 1R2A.  
 DR GO; GO:0008603; F:CAMP-dependent protein kinase regulator act. . . ; IEA.  
 DR GO; GO:0007165; P:signal transduction; IEA.  
 DR InterPro; IPR000048; IQ region.  
 DR InterPro; IPR003117; RIIA.  
 DR Pfam; PF00612; IQ; 1.  
 DR Pfam; PF02197; RIIA; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00394; RIIA; 1.  
 DR PROSITE; PS50096; IQ; 1.  
 SQ SEQUENCE 212 AA; 24304 MW; 7DCB36B0366D4CE4 CRC64;

Alignment Scores:  
 Pred. No.: 4,04e-50 Length: 212  
 Score: 684.50 Matches: 141  
 Percent Similarity: 69.34% Conservative: 6  
 Best Local Similarity: 66.51% Mismatches: 4  
 Query Match: 41.71% Indels: 61  
 DB: 2 Gaps: 2

AF334735 (1-954) x Q9XT29 (1-212)  
 QY 136 ATGTCGATTCCTCATCCACACCCACTACCGAATTCACAGGATTTGGAAATCTTCTT 195  
 DB 1 MetSerIleProPheSerAsnThrHisCysArgIleProGlnGlyPheGlyAsnLeu 20  
 QY 196 GAAGCGCTGACACCGAGATTCGAGAGACCAACCGACCAATATACAGCTTTTTCAGCA 255  
 DB 21 GluGlyLeuThrArgGluIleLeuArgGluGlnProAspAsnIleProAlaPheAlaAla 40  
 QY 256 GCCTATTTTGAAGACCTCTAGAGAAAAGAGAAAACCACTTTGATCCAGCAGATGG 315  
 DB 41 AlaTyrPheGlnSerLeuLeuGluLysArgGluLysThrAsnPheAspProAlaGluTrp 60  
 QY 316 GGGAGTAGGTAGAGACCGCTCTTATAACAATCATGCTTCGAG----- 360  
 DB 61 GlySerLysIleAspAspArgPheTyrAsnAsnHisAlaPheGluLysGlnGluProPro 80  
 QY 360 ----- 360  
 DB 81 GluLysCysAspProLysGlnGluLysSerGlnIleSerAlaLysGluGluThrPro 100  
 QY 360 ----- 360  
 DB 101 ValThrIleLeuLysGlnGluProProGluLysCysAspProLysGlnGluLysSerGln 120  
 QY 361 -----GAGCAAGAACCACTTGAGAAA 381  
 DB 121 ThrSerGlyThrGluGluGluThrProValThrIleLeuLysGlnGluProProGluLys 140  
 QY 382 AGTGATCTTAACAAGAGAGCTCTCAGATATCTCGGAGGAGGAGAGACATCAGTCACC 441  
 DB 141 CysAspProLysGlnGluLysSerGlnIleSerGlyLysGluGluThrProValThr 160



## Alignment Scores:

Pred. No.: 3 32e-40 Length: 146  
 Score: 569.00 Matches: 109  
 Percent Similarity: 86.67% Conservativeness: 21  
 Best Local Similarity: 72.67% Mismatches: 16  
 Query Match: 34.67% Indels: 3  
 DB: 1 Gaps: 3

AF334735 (1-954) x SP17\_RABIT (1-146)

QY 136 ATGTCGATTCCTTCCACACCCACTACCGAATTCACAGGATTTGGGAATCTTCT 195  
 DB 1 MetSerIleProPheSerAsnThrHisTyArgIleProGlnGlyPheGlyAsnLeuLeu 20  
 QY 196 GAAGGGCTGACACGGGAGATTCTGAGAGACCAACGGCAATATATACCGCTTTTCAGCA 255  
 DB 21 GluGlyLeuThrArgGluIleLeuArgGluGlnProAspAsnIleProAlaPheAlaAla 40  
 QY 256 GCCTATTCTTGTAGAGCCTTCTAGAGAAAAGAGAGAAAACCAACTTTGTATCCAGCAGAAATGG 315  
 DB 41 AlaTyPheGluAsnLeuLeuGluLysArgGluLysThrAsnPheAspProAlaGluTrp 60  
 QY 316 GGGAGTAGGTAGAGACCGCTTCTATACAAATCATGCTATTCGAGGAGCAAGAACCCCT 375  
 DB 61 GlyAlaLysValAspAspArgPheTyAsnAsnHisAlaPheGlnGluHisGlu---Ser 79  
 QY 376 GAGAAAGTGTCTTAACAAGAGTCTCAGATATCTGGGAAGGAGGAGAGACATCA 435  
 DB 80 GluLysCysGluAlaGluLysSerGlnSerValThr-----GluGluGluThrPro 97  
 QY 436 GTCACCATCTTAGACTCTTCTCAGGAAGATAAGGAAAAGAGAGGTTGCTCTGTCAAA 495  
 DB 98 ValLeuThrIleAsp---SerGluAspAspLysAspLysGluGluLysAlaAlaLys 116  
 QY 496 ATCCAGCTGCTTCGGGGACACATAGCCAGAGAGGAGGCAAGAAAATGAAAACAAAT 555  
 DB 117 IleGlnAlaAlaPheArgGlyHisLeuAlaArgGluAspValLysLysIleArgThrAsn 136  
 QY 556 AGTCTTCAAAATGAGGAAAAGAGGAAAC 585  
 DB 137 LysAlaGluGluGluThrGluGluAsnAsn 146

## RESULT 8

Q921K2 PRELIMINARY; PRT; 148 AA.  
 AC Q921K2  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Sp17 protein.  
 GN Name=sp17;  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Frayne J., J. A., Hall L.;  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ131888; CAA10524.1; -  
 DR HSSP; P12367; 1R2A.  
 DR GO; GO:0008603; F:GAMP-dependent protein kinase regulator act. . . ; IEA.  
 DR GO; GO:0007165; P:signal transduction; IEA.  
 DR InterPro; IPR000048; IQ region.  
 DR InterPro; IPR003117; RIIa.  
 DR Pfam; PF00612; IQ; 1.  
 DR Pfam; PF02197; RIIa; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00394; RIIa; 1.  
 DR PROSITE; PS00096; IQ; 1.  
 DR PROSITE; PS00096; IQ; 1.  
 SQ SEQUENCE 148 AA; 17096 MW; 5E0BC9D4C86F0585 CRC64;

## Alignment Scores:

Pred. No.: 1.46e-39 Length: 148  
 Score: 561.50 Matches: 109  
 Percent Similarity: 86.67% Conservativeness: 21  
 Best Local Similarity: 72.67% Mismatches: 17  
 Query Match: 34.22% Indels: 3  
 DB: 2 Gaps: 2

AF334735 (1-954) x Q921K2 (1-148)

QY 136 ATGTCGATTCCTTCCACACCCACTACCGAATTCACAGGATTTGGGAATCTTCT 195  
 DB 1 MetSerIleProPheSerAsnThrHisTyArgIleProGlnGlyPheGlyAsnLeuLeu 20  
 QY 196 GAAGGGCTGACACGGGAGATTCTGAGAGCAACCGGCAATATATACCGCTTTTCAGCA 255  
 DB 21 GluGlyLeuThrArgGluIleLeuArgGluGlnProAspAsnIleProAlaPheAlaAla 40  
 QY 256 GCCTATTCTTGTAGAGCCTTCTAGAGAAAAGAGAGAAAACCAACTTTGTATCCAGCAGAAATGG 315  
 DB 41 AlaTyPheGluAsnLeuLeuGluLysArgGluLysThrSerPheAspProAlaGluTrp 60  
 QY 316 GGGAGTAGGTAGAGACCGCTTCTATACAAATCATGCTATTCGAGGAGCAAGAACCCCT 375  
 DB 61 GlyAlaLysValAspAspArgPheTyAsnAsnHisAlaPheLysAspProGluGlnAla 80  
 QY 376 GAGAAAGTGTCTTAACAAGAGTCTCAGATATCTGGGAAGGAGGAGAGACATCA 435  
 DB 81 GluLysCysGlu-----GlnGluIleAlaLysAlaSerGlyArg---GluGluThrPro 97  
 QY 436 GTCACCATCTTAGACTCTTCTCAGGAAGATAAGGAAAAGAGAGGTTGCTCTGTCAAA 495  
 DB 98 ValThrProPheGluGluSerThrGluGluLysGluGluGluAlaAlaValLys 117  
 QY 496 ATCCAGCTGCTTCGGGGACACATAGCCAGAGAGGAGGCAAGAAAATGAAAACAAAT 555  
 DB 118 IleGlnSerAlaPheArgGlyHisValAlaArgGluGluValLysLysMetLysSerAsp 137  
 QY 556 AGTCTTCAAAATGAGGAAAAGAGGAAAC 585  
 DB 138 LysSerGluAsnValLysGluGluGluAsn 147

## RESULT 9

SP17\_MOUSE  
 ID SP17\_MOUSE STANDARD; PRT; 149 AA.  
 AC Q62252;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Sperm surface protein Sp17 (Sperm autoantigenic protein 17).  
 GN Name=Sp17; Synonyms=Sp17;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Testis;  
 RX MEDLINE=96039129; PubMed=7578682;  
 RA Kong M., Richardson R.T., Widgren E.B., O'Rand M.G.;  
 RT "Sequence and localization of the mouse sperm autoantigenic protein,  
 Sp17";  
 RL Biol. Reprod. 53:579-590 (1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler N.G.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,



RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S.C., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Murzy D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RA "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RL -!- FUNCTION: Sperm surface zona pellucida binding protein. Helps to  
 CC bind spermatozoa to the zona pellucida with high affinity. Might  
 CC function in binding zona pellucida and carbohydrates (By  
 CC similarity).  
 CC -!- SUBUNIT: Homodimer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Membrane-associated (Potential).  
 CC -!- TISSUE SPECIFICITY: Testis- and sperm-specific.  
 CC -!- SIMILARITY: Contains 1 IQ domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z46299; CAAB6455.1; -;  
 CC EMBL; BC059727; AAH59727.1; -;  
 CC F1R; S49527; S49527.  
 CC HSSP; P12367; 1R2A.  
 CC MGD; MGI:1333778; Spal17.  
 CC InterPro; IPR000048; IQ region.  
 CC InterPro; IPR003117; RIIa.  
 CC Pfam; PF00612; IQ; 1.  
 CC SMART; SM00197; RIIa; 1.  
 CC SMART; SM00394; RIIa; 1.  
 CC PROSITE; PS50096; IQ; 1.  
 KW Membrane.  
 FT DOMAIN 112 141 IQ.  
 SQ SEQUENCE 149 AA; 17296 MW; C7E05D111D6AF0DC CRC64;  
 Alignment Scores:  
 Pred. No.: 2,93e-39 Length: 149  
 Score: 558.00 Matches: 107  
 Percent Similarity: 86.49% Conservative: 21  
 Best Local Similarity: 72.30% Mismatches: 18  
 Query Match: 34.00% Indels: 2  
 DB: 1 Gaps: 1  
 AF334735 (1-954) x SP17\_MOUSE (1-149)  
 QY 136 ATGTGCGATTCCTTCCCAACCCCACTACCGAATCCCAAGGATTTGGGAATCTTCTT 195  
 Db 1 MetSerIleProPheSerAsnThrHisTyrArgIleProGlnGlyPheGlyAsnLeu 20  
 QY 196 GAAGGCTCACCGCAGATTCGAGAGACCAACCGGACATATACCACTTTTTCAGCA 255  
 Db 21 GluGlyLeuThrArgGluIleLeuArgGluGlnProAspAsnIleProAlaPheAla 40  
 QY 256 GCCTATTTTGGAGCGCTCTAGAGAAAGAGAGAAACCACTTGTATCCAGCAGATCG 315  
 Db 41 AlaTyrPheGluAsnLeuLeuGluIleAspGluIleThrSerPheAspProAlaGluIle 60  
 QY 316 GGGAGTAAGGTAGAGACCGCTTCTATAACAATCATGCTATCGAGGAGCAAGAACCACT 375  
 Db 61 GlyAlaIleValGluAspArgPheTyrAsnAsnHisAlaPheIleGluGlnVal 80  
 QY 376 GAGAAAGTGATCTCTAAACAAGAGATCTCAGATATCTGGGAAGGAGGAGACATCA 435  
 Db 136 ATGTGCGATTCCTTCCCAACCCCACTACCGAATTCACAGGATTTGGGAATCTTCTT 195

Db 81 GluIysCysGlu-----GlnGluLeuAlaIysSerSerGlyArgGluGluThrProVal 98  
 QY 436 GTCACCATCTTAGACTCTTCTGAGGAAGATAGAGAAAAGAGAGGTTCTGCTGTCAAA 495  
 Db 99 ThrProPheGluGluSerThrGluGluArgGluGluGluGluAlaAlaLeuIys 118  
 QY 496 ATCCAGCTGCTCTCCGGGACACATAGCAGAGAGAGGCGCAAGAAATGAAAAAAT 555  
 Db 119 IleGlnSerLeuPheArgGlyHisValAlaArgGluGluValIysLysMetLysSerAsp 138  
 QY 556 AGTCTTCAAAATGAGAAAAGAG 579  
 Db 139 LysAsnGluAsnLeuIysGluGlu 146  
 RESULT 10  
 SP17\_MACEU  
 ID SP17\_MACEU STANDARD; PRT; 153 AA.  
 AC O62770;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Sperm surface protein Sp17 (Sperm autoantigenic protein 17).  
 GN Name=SPAL17; Synonyms=SP17;  
 OS Macropus eugenii (Tamar wallaby).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.  
 CC NCBI TaxID=9315;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RL Wen Y., O'Rand M.G.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Sperm surface zona pellucida binding protein. Helps to  
 CC bind spermatozoa to the zona pellucida with high affinity. Might  
 CC function in binding zona pellucida and carbohydrates (By  
 CC similarity).  
 CC -!- SUBUNIT: Homodimer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Membrane-associated (Potential).  
 CC -!- TISSUE SPECIFICITY: Testis- and sperm-specific (By similarity).  
 CC -!- SIMILARITY: Contains 1 IQ domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF054289; AAC08024.1; -;  
 CC HSSP; P12367; 1R2A.  
 CC InterPro; IPR000048; IQ region.  
 CC InterPro; IPR003117; RIIa.  
 CC Pfam; PF00612; IQ; 1.  
 CC Pfam; PF02197; RIIa; 1.  
 CC SMART; SM00015; IQ; 1.  
 CC SMART; SM00394; RIIa; 1.  
 CC PROSITE; PS50096; IQ; 1.  
 KW Membrane.  
 FT DOMAIN 122 151 IQ.  
 SQ SEQUENCE 153 AA; 17327 MW; AD3A7C6B4D4E1B65 CRC64;  
 Alignment Scores:  
 Pred. No.: 4.69e-38 Length: 153  
 Score: 544.00 Matches: 104  
 Percent Similarity: 81.05% Conservative: 20  
 Best Local Similarity: 67.97% Mismatches: 21  
 Query Match: 33.15% Indels: 8  
 DB: 1 Gaps: 1  
 AF334735 (1-954) x SP17\_MACEU (1-153)



```

RL Biochem. J. 357:25-31 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Gerhard D.S.; to the EMBL/GenBank/DBJ databases.
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC075533; AAH75533.1; -.
DR GO; GO:0008603; F:camp-dependent protein kinase regulator act. . . IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR003117; RIIa.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF02197; RIIa; 1.
DR SMART; SM00015; IQ; 2.
DR SMART; SM00394; RIIa; 1.
DR PROSITE; PS50096; IQ; 2.
SQ SEQUENCE 495 AA; 56219 MW; 4BB57410B790912A CRC64;
Alignment Scores:
Pred. No.: 1.52e-22 Length: 495
Score: 364.00 Matches: 80
Percent Similarity: 58.42% Conservative: 31
Best Local Similarity: 42.11% Mismatches: 35
Query Match: 22.18% Indels: 44
DB: 2 Gaps: 5
AF334735 (1-954) x Q6DIJ9 (1-495)
QY 136 ATGTCGATTCATTCCTCCACACCCACTACCGAATTCACAGGATTTGGGATCTTCTT 195
Db 1 MetSerIleProPheSerAsnThrHisTyrArgIleProArgGlyPheAlaAsnLeu 20
QY 196 GAAGGCGCTGACGCGGAGATTCGTGAGAGACCAACCGCAATATACACGCTTTTCGACGA 255
Db 21 GluGlyLeuThrArgGluValLeuArgHisGlnProLysAspIleProLeuPheGlyAla 40
QY 256 GCCTATTTTGAGAGCCTTCTAGAGAAAAGAGAAAACCAACTTTTGATCCAGCAATGG 315
Db 41 LysTyrPheSerGlnLeuGlnArgGlnArgGlnAspThrAspPheAspProAlaGlnT 60
QY 316 GCGAGTAGGTAGAGACCGCTTCTATACATCATCATGATTCGAGGAGCAGAACACCT 375
Db 61 GlyAlaAlaLeuGluAspArgPheTyrAsnAsnTyrAspPheGlnHieThrGluLeuLys 80
QY 376 GAGAAAAGTGATCCTCTAAACAAGAGAGTCTCAGATATCTCGGAGGAGGAGAGACATCA 435
Db 81 LysTyrThr-----SerGluProSerGluLeuLys-----LysSerLysArgLysSer 96
QY 436 GTCCACATCTTAGACTCTTCTAG-----GTCTCTAG-----GAAGAT 465

```

```
Db 97 PheSerMetLeuAspAsnSerAspThrTyrProLeuSerProSerThrSerLeuGluGln 116
      :::::|||||:::|||||
QY 466 AAGGAA-----AAAGAGAGGTTCCTGCTGCTCAAAATCCAACTGCC 507
      |||||
Db 117 LysGluIleGlnThrAspMetArgLysMetGlnAlaAlaThrThrIleGlnAlaAla 136
      |||||
QY 508 TTCGGGGGACATACCCAGA----- 528
      |||||
Db 137 PheArgGlyTyrAsnValArgHisValLysIleHisGlnGluSerPheValAspAsn 156
      |||||
QY 529 -----GAGGAGGCAAAAGAAATGAAAAAAT 555
      :::::|||||:::|||||
Db 157 TyrMetLysProGluProGluHisGluAspThrAspGlnAlaLysGluThrLysCysAsp 176
      :::::|||||:::|||||
QY 556 AGCTTCAAAATGAGGAAAGAGAAAC 585
      :::::|||||:::|||||
Db 177 LeuMetGluAsnGluAspCysGluPheAsn 186
      |||||

RESULT 14
Q20199 PRELIMINARY; PRT; 189 AA.
AC Q20199;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein F39H12.3.
GN Name=F39H12.3; ORFNames=F39H12.3;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG WormBase Consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Minx M., Wohldmann P.;
RT "The sequence of C. elegans cosmid F39H12.";
RL Submitted (JAN-1996) to the EMBL/GenBank/DBSJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBSJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBSJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RG WormBase Consortium;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBSJ databases.
DR EMBL; U42847; AA883605.1; -.
DR FIR; T29159; T29159.
DR WormBase; WBGene00018214; F39H12.3.
DR WormPep; F39H12.3; CB07210.
DR GO; GO:0008603; F:camp-dependent protein kinase regulator act. . .; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR003117; RIIa.
DR Pfam; PF00612; IQ; 3.
DR Pfam; PF02197; RIIa; 1.
DR SMART; SM00015; IQ; 2.
DR SMART; SM00394; RIIa; 1.
```

```
DR PROSITE; PS50096; IQ; 2.
KW Hypothetical protein.
SQ SEQUENCE 189 AA; 21625 MW; 0C640476338DE493 CRC64;

Alignment Scores:
Pred. No.: 0.00825 Length: 189
Score: 134.00 Matches: 47
Percent Similarity: 39.25% Conservative: 26
Best Local Similarity: 25.27% Mismatches: 44
Query Match: 8.17% Indels: 69
DB: 2 Gaps: 7

AF334735 (1-954) x Q20199 (1-189)
QY 169 ATTCACACAGATTGGGAATCTTTGAAGGGCTGACACGGAGATCTCTGAGAGACAA 228
      ::|||
Db 9 ValProHisAspLeuArgProIleLeuAlaLeuAlaArgGluValLeuArgSerGln 28
      ::|||
QY 229 CCGGACAATATACCAGCTTTTTCAGACGACCTATTTTGAGAGCCTTCTAGAG---AAAAAGA 285
      |||
Db 29 ProSerAspValAlaGluPheGlyHisMetPhePheAspGluTyrLeuLysHisArgArg 48
      |||
QY 286 GAGAAACCAACTTT-----GATCCAGACAGAAATGGGGAGTAAGGTAGAGACGCGTTC 339
      |||
Db 49 GluAsnArgAsnIleLeuLysAspProAla----- 58
      |||
QY 340 TATAACAATCATGATTCGAGGAGCAAGACCACTGAGAAAGTGTATCTTAACAAGAA 399
      |||
Db 59 -----AlaTyrGlu-----ValPheArgAlaAspLeuGlnLysLys 70
      |||
QY 400 GAGTCTCAGATATCTGGGAAGGAGAGACATCAGTCACCATCTTAGACTCTTCTCGAG 459
      ::|||
Db 71 PheAlaGluVal-----Glu 75
      |||
QY 460 GAAGATAAGGAAAAAGAGAGTTGCTGCTCTCAAAATCCAAAGTGCCTTCGGGGACAC 519
      ::|||
Db 76 ArgProAlaSerProMetAspThrAlaAlaThrLysIleGlnAlaAlaPheLysGlyHis 95
      |||
QY 520 ATAGCCAGAGAGGAGGCAAGAAATG-----AAA 549
      ::|||
Db 96 LeuValArgAlaHisProGluLysTyrGlyMetSerThrArgThrSerSerSerGluLys 115
      |||
QY 550 ACAATAGTCTTCAAAATGAGAAAAAGAGAGAAA----- 583
      ::|||
Db 116 LeuAspSerAlaAsnAsn-LysLysAspGlnLysArgHisSerValGlyGlyTyrThrI 135
      |||
QY 594 -----ACAACTGAGGACACTGGTTT 603
      |||
Db 135 eAspValAspThrProGluAspArgAlaAlaThrLysIleGlnSerGluIleArgGlyPh 155
      |||
QY 604 TACTCCAGGAAACAT 619
      ::|||
Db 155 eLeuThrArgLysHis 160
      |||

RESULT 15
Q6NW93 PRELIMINARY; PRT; 422 AA.
AC Q6NW93;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein zgc:85886.
GN ORFNames=zgc:85886;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
```

RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Villalon D.K., Muzny K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Richardson D., Worley K.C., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Whiting M., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Blakesley R.W., Touchman J.W., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywiński M.I., Skalkala U., Schmutz J., Myers R.M., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Strausberg R.;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC067676; AAH67676.1; -.  
 DR ZFIN; ZDB-GENE-040426-2427; zgc:85886.  
 DR GO; GO:0005952; C:cAMP-dependent protein kinase complex; IEA.  
 DR GO; GO:0008603; F:cAMP-dependent protein kinase regulator act. .; IEA.  
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
 DR GO; GO:0007165; P:signal transduction; IEA.  
 DR InterPro; IPR002373; cAMP kin.  
 DR InterPro; IPR000595; CNMP binding.  
 DR InterPro; IPR003117; Riia.  
 DR Pfam; PF00027; CNMP binding; 2.  
 DR Pfam; PF02197; Riia; 1.  
 DR PRINTS; PR00103; CAMPKINASE.  
 DR SMART; SM00100; CNMP; 2.  
 DR SMART; SM00394; Riia; 1.  
 DR PROSITE; PS00888; CNMP\_BINDING\_1; 2.  
 DR PROSITE; PS00889; CNMP\_BINDING\_2; 2.  
 DR PROSITE; PS50042; CNMP\_BINDING\_3; 2.  
 KW Hypothetical protein.  
 SQ SEQUENCE 422 AA; 48016 MW; 4BAFB0935535D5DC CRC64;

Alignment Scores:  
 Pred. No.: 0.0159 Length: 422  
 Score: 131.00 Matches: 31  
 Percent Similarity: 50.49% Conservative: 21  
 Best Local Similarity: 30.10% Mismatches: 39  
 Query Match: 7.98% Indels: 12  
 DB: 2 Gaps: 3

AF334735 (1-954) x Q6NW93 (1-422)

QY 136 ATGTGATTCATTCCTCCACACCCACTACCGAATTCACAGGATTTGGGATCTTCTT 195  
 |||||:|||||  
 Db 1 MetSerValPro-----ValProAlaGlyLeuProGluLeuLeu 13

QY 196 GAAGGCTCACCGCGAGATCTTGAGAGACCAACCGCAATATACCAGCTTTTCAGCA 255  
 |||||:|||||  
 Db 14 ArgGlyPheThrLeuGluValLeuArgGlnProAlaAspLeuLeuGluPheSerVal 33

QY 256 GCCTATTTTGAGAGCTTCTAGAGAAAAGAGAGAAAACC-----AACTTTGAT 303  
 |||||:|||||  
 Db 34 ArgTyrPheThrGlyLeuArgPheThrArgSerSerThrArgSerAlaSerArgSer 53

QY 304 CCAGCAGAATGGGGAGTAGGTAGAA---GACCGCTTCTATAACAATCATCGATTCGAG 360  
 |||||:|||||  
 Db 54 ProValThrArgGlyGluAlaPheAspArgAspSerAlaHisThrAspSerValGlu 73

QY 361 GAGCAGAACCACTCAGAAAAGTGTCTCTAAACAAGAGAGTCTCAGATATCTGGGAG 420  
 |||||:|||||  
 Db 74 GluGluGlySerProGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 93

QY 421 GAGGAAGAG 429  
 |||||:|:  
 Db 94 GluGluAsp 96

RESULT 16  
 Q25114 PRELIMINARY; PRT; 368 AA.  
 AC Q25114;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Regulatory subunit of cAMP-dependent histone kinase.  
 OS Hemacentrotus pulcherrimus (Sea urchin).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinoidea; Euechinozoa; Echinacea; Echinozoa; Strongylocentrotidae;  
 OC Hemacentrotus.  
 OX NCBI\_TaxID=7650;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97210759; PubMed=9057823;  
 RA Hoshino K., Nomura K., Suzuki N.;  
 RT "Cyclic-AMP-dependent activation of an inter-phylum hybrid histone-  
 RT kinase complex reconstituted from sea urchin sperm-regulatory subunits  
 and bovine heart catalytic subunits.";  
 RL Eur. J. Biochem. 243:612-623(1997).  
 DR EMBL; D83379; BAA11899.1; -.  
 DR HSP; P12369; 1CX4.  
 DR GO; GO:0005952; C:cAMP-dependent protein kinase complex; IEA.  
 DR GO; GO:0008603; F:cAMP-dependent protein kinase regulator act. .; IEA.  
 DR GO; GO:0006301; P:kinase activity; IEA.  
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
 DR GO; GO:0007165; P:signal transduction; IEA.  
 DR InterPro; IPR002373; cAMP kin.  
 DR InterPro; IPR000595; CNMP binding.  
 DR InterPro; IPR003117; Riia.  
 DR Pfam; PF00027; CNMP binding; 2.  
 DR Pfam; PF02197; Riia; 1.  
 DR PRINTS; PR00103; CAMPKINASE.  
 DR SMART; SM00100; CNMP; 2.  
 DR SMART; SM00394; Riia; 1.  
 DR PROSITE; PS00888; CNMP\_BINDING\_1; 1.  
 DR PROSITE; PS00889; CNMP\_BINDING\_2; 2.  
 DR PROSITE; PS50042; CNMP\_BINDING\_3; 2.  
 KW Kinase.  
 SQ SEQUENCE 368 AA; 41679 MW; DF596699B3579186 CRC64;

Alignment Scores:  
 Pred. No.: 0.0847 Length: 368  
 Score: 122.50 Matches: 49  
 Percent Similarity: 47.09% Conservative: 40  
 Best Local Similarity: 25.93% Mismatches: 67  
 Query Match: 7.46% Indels: 33  
 DB: 2 Gaps: 7

AF334735 (1-954) x Q25114 (1-368)

QY 160 CACTACCGAATTCACAGGATTTGGGATCTTCTTGAAGGCTGACCGAGATTCG 219  
 :|||:|||||:|:  
 Db 2 AsnPheGluIleProGluGlyLeuThrAspLeuLeuGlnAspPheThrValAlaValLeu 21

QY 220 AGAGAGCAACCGCAATATACCAGCTTTTGAGAGGCTTATTTGAGAGCCTTCTAGAG 279  
 |||||:|||||  
 Db 22 ArgGluIleProSerAspLeuValGluPheAlaSerSerTyrPheAlaIleLeuGlnGlu 41

QY 280 -----AAAAGAGAGAAAACCACTTTGATTCGAGAGATGGGG 318  
 |||||:|||||  
 Db 42 AsnAsnIleSerLeuGlyIleGlyIleArgGlyValThrPheThrAlaProGluAspAlaGlu 61

QY 319 AGTAGGTAGAGACCGC-----TTCTATACAAATCATGCTTCGAGGAGCAA 366  
 |||||:|||||  
 Db 62 SerAspIleAspAspGluProGluLeuProGluLeuProGluLeuProGluLeuProGluLeu 81

QY 367 GAACACCTCGAGAAAGTGCATCTAAACAAGAGAGTCTCAGATATCTGGGAAGGAGAA 426  
 Db 82 ValCysAlaGluValTyrAspProGluAlaAsp-----AsnAspThr 95  
 QY 427 GAGACATCAGTCACCATCTAGACTCTTCTGAGAGAGTAAGAAAGAAAGAGGTGCT 486  
 Db 96 AspThrGlnLysLeuValTyrProLysSerAspGluGlnArgAlaArg---LeuThrGlu 114  
 QY 487 GCTGTCAAAATCCAGTCCCTTCGGGGACACATAGCCAGAGAGGAGCAAGAAATG 546  
 Db 115 AlaValLysAsnIleLeuLeuPheArg-----AlaLeuAspGlu---LysGlnMet 130  
 QY 547 AAAACAATAGTCTTCAAAATCAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 606  
 Db 131 -GlnGluValIleAspAlaMetPheGluLysLysThrThrProGlyAspHis-ValIleA 150  
 QY 607 CTCAGGAAACATGAAATATCAAAATCCATCAACCTCTTATTAATGTCATTTCTTC 666  
 Db 150 spGlnGlyAspAspGlyAspAsn-----PheTyrValI 161  
 QY 667 CTGAGGAGGAGATTTGATGTT 689  
 Db 161 leAspArgGlyGluTyrAspIle 168

## RESULT 17

KAPR\_STRPU STANDARD; PRT; 369 AA.  
 AC Q26619;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE CAMP-dependent protein kinase type II regulatory subunit.  
 OS Strongylocentrotus purpuratus (Purple sea urchin).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinoidea; Euechinoidea; Echinacea; Echinozoa; Strongylocentrotidae;  
 OC Strongylocentrotus.  
 OX NCBI\_TaxID=7668;  
 RN [1]\_SIMILARITY: Belongs to the CAMP-dependent kinase regulatory chain family.  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Rankin T.L., Bangott L.J.;  
 RT "Regulatory subunit type II of cyclic AMP-dependent protein kinase localized in sea urchin spermatozoa."  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: Tetramer, composed of 2 regulatory (R) and 2 catalytic (C) subunits. In the presence of CAMP it dissociates into 2 active monomeric C subunits and an R dimer (by similarity).  
 CC -1- SIMILARITY: Belongs to the CAMP-dependent kinase regulatory chain family.  
 CC -1- SIMILARITY: Contains 2 cyclic nucleotide-binding domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U19887; AAA61966.1; -.  
 CC HSP; P12369; 1CX4.  
 CC InterPro; IPR002373; CAMP\_kin.  
 CC InterPro; IPR000595; CNMP\_binding.  
 CC InterPro; IPR003117; Riia.  
 CC Pfam; PF02197; Riia; 1.  
 CC PRINTS; PR00103; CAMPKINASE.  
 CC SMART; SM00100; CNMP; 2.  
 CC SMART; SM00394; Riia; 1.  
 CC PROSITE; PS00888; CNMP\_BINDING\_1; 1.  
 CC PROSITE; PS00889; CNMP\_BINDING\_2; 2.  
 CC PROSITE; PS0042; CNMP\_BINDING\_3; 2.  
 CC CAMP-binding; Phosphorylation; Repeat.

FT DOMAIN 1 120 Dimerization and phosphorylation (By similarity).  
 FT NP\_BIND 121 238 CAMP 1.  
 FT NP\_BIND 239 365 CAMP 2.  
 FT MOD\_RES 81 81 Phosphoserine (By similarity).  
 FT BINDING 186 186 CAMP 1 (By similarity).  
 FT BINDING 195 195 CAMP 1 (By similarity).  
 FT BINDING 313 313 CAMP 2 (By similarity).  
 FT BINDING 322 322 CAMP 2 (By similarity).  
 SQ SEQUENCE 369 AA; 41788 MW; 30FCECC13D8C26A5 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.0848 Length: 369  
 Score: 122.50 Matches: 49  
 Percent Similarity: 47.09% Conservative: 40  
 Best Local Similarity: 25.93% Mismatches: 67  
 Query Match: 7.46% Indels: 3  
 DB: 1 Gaps: 7  
 AF334735 (1-954) x KAPR\_STRPU (1-369)  
 QY 160 CACTACCAATTCACAAAGATTGGGAATCTTCTTGAAGGGCTGACAGCGAGATTCTG 219  
 Db 2 AsnPheGluIleProGluGlyLeuThrAspLeuLeuGlnAspPheThrValAlaValLeu 21  
 QY 220 AGAGAGCAACCGACCAATATACCACTTTTCAGCAGCCTATTTTGAGAGCCTCTAGAG 279  
 Db 22 ArgGluLysProSerAspLeuValGluPheAlaSerSerTyrPheAlaLysLeuGlnGlu 41  
 QY 280 -----AAAAGAGAGAAACCAACTTTTGATCCAGCAGAAATGGGG 318  
 Db 42 AsnAsnIleSerLeuGlyGlyLysArgGlyValThrPheThrAlaProGluAspAlaGlu 61  
 QY 319 AGTAGGTAGAGACCGC-----TTCTATAACAATCATGCTATCGAGAGCA 366  
 Db 62 SerAspIleAspAspGluProProGluLeuProLysAsnArgTyrAlaArgLysSer 81  
 QY 367 GAACACCTCGAGAAAGTGCATCTAAACAAGAGAGTCTCAGATATCTGGGAAGGAGAA 426  
 Db 82 ValCysAlaGluLysTyrAspProGluAlaAsp-----AsnAspThr 95  
 QY 427 GAGACATCAGTCACCATCTAGACTCTTCTGAGAGAGTAAGAAAGAAAGAGGTGCT 486  
 Db 96 AspThrGlnLysLeuValTyrProLysSerAspGluGlnArgAlaArg---LeuThrGlu 114  
 QY 487 GCTGTCAAAATCCAGTCCCTTCGGGGACACATAGCCAGAGAGGAGCAAGAAATG 546  
 Db 115 AlaValLysAsnIleLeuLeuPheArg-----AlaLeuAspGlu---LysGlnMet 130  
 QY 547 AAAACAATAGTCTTCAAAATCAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 606  
 Db 131 -GlnGluValIleAspAlaMetPheGluLysLysThrThrProGlyAspHis-ValIleA 150  
 QY 607 CTCAGGAAACATGAAATATCAAAATCCATCAACCTCTTATTAATGTCATTTCTTC 666  
 Db 150 spGlnGlyAspAspGlyAspAsn-----PheTyrValI 161  
 QY 667 CTGAGGAGGAGATTTGATGTT 689  
 Db 161 leAspArgGlyGluTyrAspIle 168  
 RESULT 18  
 QYHAY3 PRELIMINARY; PRT; 221 AA.  
 AC Q9HAY3;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Testis-specific calcium-binding protein CBP86-IV.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;







```
QY 415 GGAAGGAGGAGGAGACATCAGTCACCATCTTGTAGACTCTTCTGAGAGAGATAAGGAAAA 474
Db 103 ThrAspThrAspGluAspValThrArgThrGluThrSerAspLysThrThrGlnPhe 122

QY 475 GAAGAGGTTGCTGCTGTC 492
Db 123 ProSerValTyAlaVal 128

RESULT 20
O75952 PRELIMINARY; PRT; 493 AA.
AC O75952;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fibrous sheathin II.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testes;
RX MEDLINE=21678993; PubMed=11820818; DOI=10.1006/dbio.2001.0527;
RA Naaby-Hansen S., Mandal A., Wolkowicz M.J., Sen B., Westbrook V.A.,
RA Shetty J., Coonrod S.A., Klotz K.L., Kim Y.H., Bush L.A.,
RA Flickinger C.J., Herr J.C.;
RT "CABYR, a novel calcium-binding tyrosine phosphorylation-regulated
RT fibrous sheath protein involved in capacitation.";
RL Dev. Biol. 242:236-254(2002).
DR EMBL; AF088868; AAC35373.1; -.
DR GenBank; HGNC:15569; CABYR.
DR GO; GO:0008603; F:CaMP-dependent protein kinase regulator act. . . ; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR003117; RIIa.
DR Pfam; PF02197; RIIa; 1.
DR SMART; SM00394; RIIa; 1.
SQ SEQUENCE 493 AA; 52788 MW; 0911CDA626325C27 CRC64;

Alignment Scores:
Pred. No.: 0.106 Length: 493
Score: 121.50 Matches: 35
Percent Similarity: 46.03% Conservative: 23
Best Local Similarity: 27.78% Mismatches: 43
Query Match: 7.40% Indels: 25
DB: 2 Gaps: 3

AF334735 (1-954) x O75952 (1-493)
QY 169 ATTCCACAGGATTGGGAACTCTTGAAGGCTGACACGCGAGATCTTGAGAGCAA 228
Db 10 ValProTyrGlyLeuLysThrLeuLeuGluGlyIleSerArgAlaValLeuLysThrAsn 29

QY 229 CCGGACAAATACCAAGCTTTTCAGCAGCGCTTTTCAGAGCGCTCTTAGAGAAAAGAGAG 288
Db 30 ProSerAsnIleAsnGlnPheAlaAlaTyrPheGlnLeuThrMetTyrArgGly 49

QY 289 AAAACCAACTTGATCCAGCAGATGGGGGAGTAAAGGTAGAGACCGCTTCTATACCAAT 348
Db 50 AsnThrThrMetAsp-----IleLysAspLeuValLysGlnPhe 62

QY 349 CATGCATTCGAGGACGAACACCACTGAGAAAGTGATCTTAACAGAA----- 399
Db 63 HisGlnIleLysValGluLysTrpSerGluGlyThrThrProGlnLysLysLeuGluCys 82

QY 400 -----GAGTCTCAGATA-----TCT 414
Db 83 LeuLysGluProGlyLysThrSerValGluSerLysValProThrGlnMetGluLysSer 102

QY 415 GGAAGGAGGAGAGACATCAGTCACCATCTTGTAGACTCTTCTGAGAGAGATAAGGAAAA 474
Db 103 ThrAspThrAspGluAspValThrArgThrGluThrSerAspLysThrThrGlnPhe 122
```

```
QY 475 GAAGAGGTTGCTGCTGTC 492
Db 123 ProSerValTyAlaVal 128

RESULT 21
Q9GM47 PRELIMINARY; PRT; 596 AA.
AC Q9GM47;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain parietal lobe;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB049839; BAB16725.2; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0006412; P:Protein biosynthesis; IEA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR003890; IF_eIF4G.
DR InterPro; IPR007193; Upf2.
DR Pfam; PF02854; MIF4G; 2.
DR Pfam; PF04050; Upf2; 1.
DR SMART; SM00543; MIF4G; 2.
DR Hypothetical protein.
KW SEQUENCE 596 AA; 69788 MW; F70AD5DD09E165A0 CRC64;

Alignment Scores:
Pred. No.: 0.118 Length: 596
Score: 121.00 Matches: 41
Percent Similarity: 47.89% Conservative: 27
Best Local Similarity: 28.87% Mismatches: 48
Query Match: 7.37% Indels: 26
DB: 2 Gaps: 6

AF334735 (1-954) x Q9GM47 (1-596)
QY 265 GAGACCTCTTAGAGAAAAGAGAGAAACCACTTTGATCCAGCAGATGGGGAGTAAG 324
Db 439 GluAsnLeuGluGluAspGluGlu-----GluGluGlyAlaGlu 453

QY 325 GTAGAAGACCGCTCTTATAACAATCATCGATTCGAGGAGCAAGACCACCTGAGAAAGT 384
Db 454 ThrGluGluGlnSerGlyAsnGluSerGluValAsnGluProGluGluGluGlySer 473

QY 385 GATCTAAACAAGAGAGTCTCAGATATCTGGGAGGAGGAGAGACATCAGTCCATC 444
Db 474 AsnAspAspAspGluGlyGlu-----GluGluGluGluGluAsnThrAspTyrLeu 491

QY 445 TTAGACTCTTCGAGAGAGATAAGGAAAAGAGAGGTTGCTGCTGTCTCAATCAAGCT 504
Db 492 ThrAspSerAsnLysGluAsnGluThrAspGluGluSerThrGlyValMetIleLysGly 511

QY 505 GCCTCCGGGAGCAGATA-----GCCAGAGAGAGGCAAGAAATAATGAAACAAAT--- 555
Db 512 GlyGlyLeuLysHisValProCysValGluGluGlnAlaLysArgGlnGlnGluGlu 531

QY 556 ---AGTCTCTCAAAATGAGGAAAAGAGGAAAACCAAGTCAGGACACTGGTTTACCTCCAG 612
Db 532 AlaAlaAlaGlnMetLysGluLysGluGluSer----- 542

QY 613 GAAACATGAAATAATCAATCAATCACTCTTATTAATGTCTTCTTCTCTGAGG 672
Db 543 -----ValGlnLeuHisGlnGluAlaTrpGluArgHis---HisLeuArg 556
```

```

QY 673 AACGAA 678
DB 557 LysGlu 558

RESULT 22
Q6BU41
ID Q6BU41 PRELIMINARY; PRT; 562 AA.
AC Q6BU41;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similarities with CA4297|IPF2603 Candida albicans IPF2603.
GN ORFNames=DEHA0C14828g;
OS Debaryomyces Hansenii CBS767.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
OC NCBI_TaxID=284592;
RN [1]_TaxID=284592;
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RG Genolevures;
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boissrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Ostas S., Ozier-Kalogeropoulos O.,
RA Pellenn S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swenne D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RN Nature 430:35-44(2004).
(2)
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CB382135; CAG86356.1; -
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; Bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PS0014; BROMODOMAIN 2; 1.
SQ SEQUENCE 562 AA; 64900 MW; 395B482BC6EAD44F CRC64;

Alignment Scores:
Pred. No.: 0.159 Length: 562
Score: 119.50 Matches: 40
Percent Similarity: 41.46% Conservative: 28
Best Local Similarity: 24.39% Mismatches: 70
Query Match: 7.28% Indels: 26
DB: Gaps: 4

AF334735 (1-954) x Q6BU41 (1-562)
QY 108 GGTTCATGAGCGAGTCTTTACCAAGAAGATGTC-----GATTCCATTCCTCAACAC 158
DB 185 GlyGluValGlnLysAlaTyrGluGlnAspIleGluAlaGluAspAspGlnGluAsp 204
QY 159 CCATCTACCGAATTCACAGAGATTGGGAATCTCTTTGAGGGCTGACACGGGAGATCTT 218
DB 205 LysIleGlyLeuAspValGluValGluHisAspHisAsnGluLysAspIleGluAspIle 224
QY 219 GAGAGAGACACCGGACAAATATACACGCTTTTCGACAGCCTATTTTGAGAGCCTTCTAGA 278
DB 225 Glu-GluHisGluAsnAsn-----GluLeuAs 233
QY 279 GAAAGAGAGAAACCAACCACTTTTGTATCCAGAGATGGGGGAGTAGGTAGAGACCGCTT 338

```

```

DB 233 pLysGluGluThrGluAsnIleGluLeuGlnGlyAspAsnIleGluGluGlu 253
QY 339 CTATAACAATCATGCAATTCGAGGACCAAGAACCTCGAGAAAGTGTCTCTTAAACAAGA 398
DB 253 nValGluAsn-----LysAspGluAsnProAspGluProGluAspLysGluGlnG 270
QY 399 AGAGTCTCAGATATCTGGGAAGAGAGAGACATCAGTCACCATCTTTAGACTCTTCTGA 458
DB 270 uGluAspLysGluGlnLysGluGlnGluGluLeu-----G 282
QY 459 GGAACATTAAGAAAAGAGAGAGAGTGTCTCTCAAAATCCAGCTGCCTCCGGGGACA 518
DB 282 uGluAspLysGluGlnGluGlnGluGlnGluGlnGluGlnGluGlnGluGlnGluGln 302
QY 519 CATAGCCAGAGAGAGGAGCAAGAAATGAAACAAATAGTCTTCAAAATGAGGAGAAAGA 578
DB 302 nGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 322
QY 579 GGAACAAACAG 588
DB 322 uLysGluGlu 325

RESULT 23
Q9BUB1
ID Q9BUB1 PRELIMINARY; PRT; 382 AA.
AC Q9BUB1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE PKAR2A protein (protein kinase, cAMP-dependent, regulatory, type II, alpha).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
EX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmitz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
(2)
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
(3)
RP SEQUENCE FROM N.A.
RA Kalnine N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S.,
RA Koundinya M., Raphael J., Moreira D., Kelley T., LeBaer J., Lin Y.,
RA Phelan M., Farmer A.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002763; AA02763.1; -
DR EMBL; BT007225; AAP35889.1; -

```

DR HSP; P12367; 1R2A.  
 DR GO; GO:0005952; C:cAMP-dependent protein kinase complex; IEA.  
 DR GO; GO:0008603; P:cAMP-dependent protein kinase regulator act. . .; IEA.  
 DR GO; GO:0036301; P:kinase activity; IEA.  
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
 DR GO; GO:0007165; P:signal transduction; IEA.  
 DR InterPro; IPR002373; cAMP\_kin.  
 DR InterPro; IPR005955; cAMP\_binding.  
 DR InterPro; IPR003117; Riia.  
 DR Pfam; PF00027; cAMP\_binding; 2.  
 DR Pfam; PF02197; Riia; 1.  
 DR PRINTS; PR00103; CAMPKINASE.  
 DR SMART; SM00100; cAMP; 2.  
 DR SMART; SM00394; Riia; 1.  
 DR PROSITE; PS00888; cAMP\_BINDING\_1; 1.  
 DR PROSITE; PS00889; cAMP\_BINDING\_2; 2.  
 DR PROSITE; PS00442; cAMP\_BINDING\_3; 2.  
 KW Kinase.  
 SQ SEQUENCE 382 AA; 43066 MW; 87D6AC1A5CE9E300 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.188 Length: 382  
 Score: 118.50 Matches: 38  
 Percent Similarity: 41.26% Conservative: 21  
 Best Local Similarity: 26.57% Mismatches: 53  
 Query Match: 7.22% Indels: 31  
 DB: 2 Gaps: 4  
 AF334735 (1-954) x Q9BUB1 (1-382)  
 QY 157 ACCCACTACCGAATCCCAAGGATTGGGAATCTTTGAAGGCTCAGCGGAGATT 216  
 DB 2 SerHisIleGlnIleProGlyLeuThrGluLeuLeuGlnGlyThrValGluVal 21  
 QY 217 CTGAGAGACACCGGACCAATATACCAAGCTTTTGACGAGCCTATTTTGAGAGCCTCTTA 276  
 DB 22 LeuArgGlnProProAspLeuValGluPheAlaValGluTyrPheThrArgLeuArg 41  
 QY 277 GAGAAAGAGAGAAACCACTTTGATCCAGCAAGATGGGGAGTAAAGTAGAGACCGC 336  
 DB 42 GluAlaArgAlaProAlaSerValLeuProAla-----AlaThrProArg 56  
 QY 337 TTCTATAACATCATGATTCAGGAGCAAGAACCACTGAGAAAGTGATCCTAAACAA 396  
 DB 57 GlnSerLeuGlyHisProProGluProGlyProAspArgValAlaAspAlaGly 76  
 QY 397 GAAGAGTCTCAGATATCTGGGAGGAGGAGGAGACATCAGTCACCATCTTAGACTCTTCT 456  
 DB 77 Asp-----SergluSer 80  
 QY 457 GAGGAAGATAGGAAAGAGAGGTTGCTGCTGTCATAAATCCAAAGTCCTTCGGGGA 516  
 DB 81 GluGluAspGluAspLeuGlu-----ValProValProSerArgPheAsnArg 96  
 QY 517 CACATAGCAGAGAGGAGGCAAGAAATGAAACAAATAGTCTTCAAAATGAGGAAAAA 576  
 DB 97 ArgValSer-----ValCysAlaGluThrTyrAsnProAspGluGlu 110  
 QY 577 GAGGAAAAAC 585  
 DB 111 GluGluAsp 113  
 RESULT 24  
 ID KAP2 HUMAN  
 AC KAP2 HUMAN STANDARD; PRT; 403 AA.  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE cAMP-dependent protein kinase type II-alpha regulatory subunit.  
 GN Name=PRKAR2A; Synonyms=PKR2, PRKAR2;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=89211413; PubMed=2540040; DOI=10.1016/0014-5793(89)80253-4;  
 RA Owen O., Myklebust F., Scott J.D., Hansson V., Jahnson T.,  
 RT "Human testis cDNA for the regulatory subunit RII alpha of cAMP-  
 RT dependent protein kinase encodes an alternate amino-terminal region.";  
 RL FEBS Lett. 246:57-64(1989).  
 [2]  
 RP SEQUENCE OF 1-86 FROM N.A.  
 RX MEDLINE=97157082; PubMed=9003463; DOI=10.1016/S0167-4781(96)00152-2;  
 RA Foss K.B., Solberg R., Simard J., Myklebust F., Hansson V.,  
 RT "Molecular cloning, upstream sequence and promoter studies of the  
 RT human gene for the regulatory subunit RII alpha of cAMP-dependent  
 RT protein kinase.";  
 RL Biochim. Biophys. Acta 1350:98-108(1997).  
 CC -!- FUNCTION: Type II regulatory chains mediate membrane association  
 CC by binding to anchoring proteins, including the MAP2 kinase.  
 CC -!- SUBUNIT: The inactive form of the enzyme is composed of two  
 CC regulatory chains and two catalytic chains. Activation by cAMP  
 CC produces two active catalytic monomers and a regulatory dimer that  
 CC binds four cAMP molecules.  
 CC -!- TISSUE SPECIFICITY: Four types of regulatory chains are found: I-  
 CC alpha, I-beta, II-alpha, and II-beta. Their expression varies  
 CC among tissues and is in some cases constitutive and in others  
 CC inducible.  
 CC -!- PTM: Phosphorylated by the activated catalytic chain.  
 CC -!- SIMILARITY: Belongs to the cAMP-dependent kinase regulatory chain  
 CC family.  
 CC -!- SIMILARITY: Contains 2 cyclic nucleotide-binding domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/  
 CC or send an email to license@ebi.ac.uk).  
 CC  
 CC EMBL; X14968; CAA33094.1; -.  
 CC EMBL; X99455; CAA67817.1; -.  
 CC PIR; S03885; OKH2R.  
 CC HSP; P12367; 1R2A.  
 CC OGP; P13861; -.  
 CC Genew; HGNC:9391; PRKAR2A.  
 CC MIM; 176910; -.  
 CC GO; GO:0005737; C:cytoplasm; TAS.  
 CC GO; GO:0005624; C:membrane fraction; TAS.  
 CC GO; GO:0005886; C:plasma membrane; TAS.  
 CC GO; GO:0007242; P:intracellular signaling cascade; TAS.  
 CC InterPro; IPR002373; cAMP\_kin.  
 CC InterPro; IPR000595; cAMP\_binding.  
 CC InterPro; IPR003117; Riia.  
 CC Pfam; PF00027; cAMP\_binding; 2.  
 CC Pfam; PF02197; Riia; 1.  
 CC PRINTS; PR00103; CAMPKINASE.  
 CC SMART; SM00100; cAMP; 2.  
 CC SMART; SM00394; Riia; 1.  
 CC PROSITE; PS00888; cAMP\_BINDING\_1; 2.  
 CC PROSITE; PS00889; cAMP\_BINDING\_2; 2.  
 CC PROSITE; PS00442; cAMP\_BINDING\_3; 2.  
 KW Acetylation; cAMP-binding; Multigene family; Phosphorylation; Repeat.  
 FT INIT MET 0  
 FT DOMAIN 1 137 Dimerization and phosphorylation.  
 FT NP\_BIND 138 259 CAMP 1.  
 FT NP\_BIND 260 403 CAMP 2.  
 FT MOD\_RES 1 1 N-acetyls erine (Probable).  
 FT MOD\_RES 98 98 Phosphoserine (By similarity).  
 FT BINDING 207 207 CAMP 1.  
 FT BINDING 216 216 CAMP 1.



DR InterPro; IPR002373; CAMP kin.  
 DR InterPro; IPR000595; CNMP\_binding.  
 DR InterPro; IPR003117; RIIa.  
 DR Pfam; PF00027; CNMP\_binding; 2.  
 DR Pfam; PF02197; RIIa; 1.  
 DR PRINTS; PR00103; CAMPKINASE.  
 DR SMART; SM00100; CNMP; 2.  
 DR SMART; SM00394; RIIa; 1.  
 DR PROSITE; PS00888; CNMP\_BINDING\_1; 2.  
 DR PROSITE; PS00889; CNMP\_BINDING\_2; 2.  
 DR PROSITE; PS00442; CNMP\_BINDING\_3; 2.  
 DR Kinase.  
 SQ SEQUENCE 401 AA; 45540 MW; 026FEBFCFACA9C7D CRC64;

Alignment Scores:  
 Pred. No.: 0.28 Length: 401  
 Score: 116.50 Matches: 38  
 Percent Similarity: 42.66% Conservative: 23  
 Best Local Similarity: 26.57% Mismatches: 49  
 Query Match: 7.10% Indels: 33  
 DB: 2 Gaps: 5

AF334735 (1-954) x Q8K1M2 (1-401)

QY 157 ACCCACTACCGAATTCACAGGATTTGGGAATCTTCTTGAAGGGCTGACACGCGAGATT 216  
 Db 2 SerHisIleGlnIleProProGlyLeuThrGluLeuGlnGlyThrValGluVal 21  
 QY 217 CTGAGAGACAACCGGACAATATACAGCTTTTGAGCAGCCTATTGAGAGCCTCTTA 276  
 Db 22 LeuArgGlnProProAspLeuValAspPheAlaValGlnThrPheThrArgLeuArg 41  
 QY 277 GAGAAAGAGAGAAACCAACTTTGATCCAGCAGATGGGGAGTAAGGTAGAACCGC 336  
 Db 42 GluAlaArgArgGlnGluSer-----AspSer 50  
 QY 337 TTTCTATAACAATCATGCAATTCGAGGACGAAGAACCACTGAGAAAGTGCCTAAACAA 396  
 Db 51 PheIle-----AlaProProThrThrPheHisAlaGln 61  
 QY 397 GAAGAGTCTCAGATATCTGGGAGAGGAGAGACATCAGTCACCATCTTAGACTTCT 456  
 Db 62 GluSerSerGlyValProValIleGluGluAsp-----GlyGluSer 75  
 QY 457 GAGGAAGATAAGGAAAGAACAGAGTGTGCTGTCAAAATCAAGCTGCCTCCGGGA 516  
 Db 76 GluSerAspSerAspAspGluAspLeu-----GluValProIleProSerLysPheThrArg 94  
 QY 517 CACATAGCCAGAGAGGAGGACAAAGAAATGAAACAAATAGTCTTCAAAATGAGGAAAAA 576  
 Db 95 ArgValSer-----ValCysAlaGluThrPheAsnProAspGluGlu 108  
 QY 577 GAGGAAAC 585  
 Db 109 GluAspAsn 111

RESULT 27

QY Q710D7 PRELIMINARY; PRT; 498 AA.  
 AC Q710D7;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Sperm protein 13.  
 OS Vulpes vulpes (Red fox).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Vulpes.  
 OX NCBI\_TaxID=9627;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Verdier Y.;  
 RT "Selection, identification et caracterisation partielle d'antigenes

RT spermatisques de renards.";  
 RL Theais (2002), Department of Biological Sciences, Universite Nancy I,  
 RL Nancy, France.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Verdier Y., Farre G., Rouet N., Boue F.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ421969; CAD19163.1;  
 DR GO; GO:0008603; P: CAMP-dependent protein kinase regulator act. . .; IEA.  
 DR GO; GO:0007165; F: signal transduction; IEA.  
 DR InterPro; IPR003117; RIIa.  
 DR Pfam; PF02197; RIIa; 1.  
 DR SMART; SM00394; RIIa; 1.  
 SQ SEQUENCE 498 AA; 53087 MW; CFFD605645421BA32 CRC64;

Alignment Scores:  
 Pred. No.: 0.285 Length: 498  
 Score: 116.50 Matches: 34  
 Percent Similarity: 48.62% Conservative: 19  
 Best Local Similarity: 31.19% Mismatches: 37  
 Query Match: 7.10% Indels: 19  
 DB: 2 Gaps: 4

AF334735 (1-954) x Q710D7 (1-498)

QY 169 ATTCACAAGGATTTGGGAATCTTCTTGAAGGGCTGACACGCGAGATTCTGAGAGCA 228  
 Db 10 ValProTyrGlyLeuLysThrLeuLeuGluGlyValSerArgAlaIleLeuLysIleAsn 29  
 QY 229 CCGGACAAATATACCAAGCTTTTGACAGCCTTATTGAGAGCCTTCTAGAGAAAGAGAG 288  
 Db 30 ProProAsnIleThrGlnPheAlaAlaValTyrPheLysGluLeuIleValPheArgGlu 49  
 QY 289 ---AAAACCAACTTGTATCCAGCAGAAATGGGGAGTAGGTAGAGACCGCTTCTATTAAC 345  
 Db 50 GlyAsnThrSerLeuAsp-----IleLysAspLeuValLysGln 62  
 QY 346 AATCATGCTATTCGAG-----GAGCAAGAACCACTCTGAG 378  
 Db 63 PheHisGlnIleLysValGluLysTrpSerGluGlyThrThrGlnGluLysGluProGlu 82  
 QY 379 AAAAGTGATCTCTAAACAGAGAGTCTCAGATATCTGGGAGAGGAGAGACATCAGTC 438  
 Db 83 CysMetGluGluGlnValGluThrSerValValSer-----GlnGluProThrArgMet 100  
 QY 439 ACCATCTTAGACTCTCTGAGAGAGAT 465  
 Db 101 GluLysSerThrAspThrGluGluAsp 109

RESULT 28

KAP2\_BOVIN  
 ID KAP2\_BOVIN STANDARD; PRT; 400 AA.  
 AC P00515;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE CAMP-dependent protein kinase type II-alpha regulatory subunit.  
 GN Name=PRKAR2A;  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Heart muscle;  
 RX MEDLINE=8222175; PubMed=6283532;  
 RA Takio K., Smith S.B., Krebs E.G., Walsh K.A., Titani K.;  
 RT "Primary structure of the regulatory subunit of type II CAMP-dependent  
 RT protein kinase from bovine cardiac muscle.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:2544-2548 (1982).  
 RN [2]

3D-STRUCTURE MODELING.  
 MEDLINE=87157645; PubMed=3030405;  
 Weber I.T., Steitz T.A., Babis J., Taylor S.S.;  
 "Predicted structures of cAMP binding domains of type I and II  
 regulatory subunits of cAMP-dependent protein kinase.";  
 Biochemistry 26:343-351(1987).  
 CC -I- SUBUNIT: The inactive form of the enzyme is composed of two  
 regulatory chains and two catalytic chains. Activation by cAMP  
 produces two active catalytic monomers and a regulatory dimer that  
 binds four cAMP molecules.  
 CC -I- TISSUE SPECIFICITY: Four types of regulatory chains are found: I-  
 alpha, I-beta, II-alpha, and II-beta. Their expression varies  
 among tissues and is in some cases constitutive and in others  
 inducible.  
 CC -I- PTM: A second phosphorylation site has not been located.  
 CC -I- SIMILARITY: Belongs to the cAMP-dependent kinase regulatory chain  
 family.  
 CC -I- SIMILARITY: Contains 2 cyclic nucleotide-binding domains.  
 DR FIR; A06618; ORK02R.  
 DR PDB; 2APK; Model; @=1-400.  
 DR InterPro; IPR002373; cAMP\_kin.  
 DR InterPro; IPR000595; cNMP\_binding.  
 DR InterPro; IPR003117; RIIa.  
 DR Pfam; PF00027; cNMP\_binding; 2.  
 DR Pfam; PF02197; RIIa; 1.  
 DR PRINTS; PR00103; CAMPKINASE.  
 DR SMART; SM00100; cNMP; 2.  
 DR SMART; SM00394; RIIa; 1.  
 DR PROSITE; PS00888; cNMP\_BINDING\_1; 2.  
 DR PROSITE; PS00889; cNMP\_BINDING\_2; 2.  
 DR PROSITE; PS0042; cNMP\_BINDING\_3; 2.  
 KW 3D-structure; Acetylation; cAMP-binding; Direct protein sequencing;  
 KW Multigene family; Phosphorylation; Repeat.  
 FT DOMAIN 1 134 Dimerization and phosphorylation.  
 FT NP\_BIND 135 256 CAMP 1.  
 FT NP\_BIND 257 400 CAMP 2.  
 FT MOD\_RES 1 1 N-acetylserine.  
 FT MOD\_RES 95 95 Phosphoserine (by PKA).  
 FT BINDING 204 204 CAMP 1.  
 FT BINDING 213 213 CAMP 1.  
 FT BINDING 334 334 CAMP 2.  
 FT BINDING 343 343 CAMP 2.  
 FT HELIX 144 149  
 FT TURN 150 150  
 FT STRAND 154 156  
 FT TURN 158 159  
 FT STRAND 163 163  
 FT TURN 165 166  
 FT STRAND 168 168  
 FT STRAND 171 172  
 FT STRAND 174 174  
 FT STRAND 177 177  
 FT STRAND 179 184  
 FT STRAND 192 198  
 FT STRAND 202 202  
 FT TURN 205 206  
 FT STRAND 212 212  
 FT STRAND 216 217  
 FT STRAND 223 225  
 FT STRAND 228 229  
 FT HELIX 230 236  
 FT TURN 237 239  
 FT TURN 241 241  
 FT HELIX 242 263  
 FT HELIX 266 270  
 FT TURN 271 272  
 FT STRAND 277 278  
 FT TURN 280 281  
 FT STRAND 283 285  
 FT STRAND 290 291  
 FT STRAND 293 298  
 FT STRAND 304 306  
 FT STRAND 311 311

FT STRAND 316 316  
 FT STRAND 321 322  
 FT TURN 328 329  
 FT TURN 332 333  
 FT HELIX 334 336  
 FT STRAND 341 342  
 FT STRAND 346 348  
 FT STRAND 353 359  
 FT HELIX 360 367  
 FT TURN 368 369  
 FT TURN 371 371  
 FT HELIX 372 378  
 FT TURN 379 379  
 FT HELIX 380 390  
 SQ SEQUENCE 400 AA; 44962 MW; 47809DD900B244E7 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.341 Length: 400  
 Score: 115.50 Matches: 46  
 Percent Similarity: 39.58% Conservative: 30  
 Best Local Similarity: 23.96% Mismatches: 61  
 Query Match: 7.04% Indels: 55  
 DB: 1 Gaps: 8  
 AF334735 (1-954) x KAP2\_BOVIN (1-400)  
 QY 157 ACCCACTACCGAATTCACAGGATTGGGAATCTTTGAAGGCTGACGCGAGATT 216  
 Db 1 SerHisIleGlnIleProGlyLeuThrGluLeuLeuGlnGlyTyrThrValGluVal 20  
 QY 217 CTGAGAGACCAACCGACAATATACAGCTTTTGACGAGCCTATTTGAGAGCCTTCTA 276  
 Db 21 LeuArgGlnArgProAspLeuValAspPheAlaValAspTyrPheThrArgLeuArg 40  
 QY 277 GAGAAAAGAGAGAAACC-----AACCTTT 300  
 Db 41 GluAlaArgSerArgAlaSerThrProProAlaAlaProProSerGlySerGlnAspPhe 60  
 QY 301 GATCCA-----GCAGAAATGGGGAGTAAGTAGAAGAC----- 333  
 Db 61 AspProGlyAlaGlyLeuValAlaAspAlaValAlaAspSerGluSerGluAspGluGlu 80  
 QY 334 -----CGCTTC----- 339  
 Db 81 AspLeuAspValProIleProGlyArgPheAspArgArgValSerValCysAlaGluThr 100  
 QY 340 TATACAAATCATGCTTCGAGGAGCAAGAACCCACCTGAGAAAGTGATCCTAAACAAGAA 399  
 Db 101 TyrAsnProAspGluGluAspThrAspProArgValIleHisProLysThrAsp 120  
 QY 400 GAGTCTCAGATATCTGGGAAGGAGGAGACATCAGTCACCATCTTAGACTCTTCGAG 459  
 Db 121 GlnGlnArgCysArgLeuGlnGluAlaCysLysAspIleLeuLeuPhe----- 136  
 QY 460 GAAGATAAGGAAAAAGAGAGTCTGCTGTCAAATCCAAAGTCCCTCCGGGACAC 519  
 Db 137 LysAsnLeuAspProGluGlnLeuSerGlnVal---LeuAspAlaMetPhe----- 152  
 QY 520 ATAGCCAGAGGAGGCAAGAAATAGTCTCAAATAGTCTTCAAATGAGGAAAAAGAG 579  
 Db 153 -----GluArgThrValLysValAspGluHisValIleAspGlnGlyAspGly 169  
 QY 580 GAAACAAGTGAGGACACTGGTTTACCTCCAGGAA 615  
 Db 170 AspAsn-----PheTyrValIleGlu 176  
 RESULT 29  
 Q9T069  
 ID Q9T069 PRELIMINARY; PRT; 532 AA.  
 AC Q9T069;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)

```

DE Hypothetical protein AT4g37820 (Hypothetical protein
DE AT4g37820/T28119.100).
GN Name=AT4g37820; Synonyms=At4g37820/T28119.100;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroside II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Robben J., Grymonprez B., Volckaert G., Bancroft I.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RA Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Robben J., Grymonprez B., Volckaert G., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinzaki K.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL035709; CAB38930.1; -
DR EMBL; AL161592; CAB80447.1; -
DR EMBL; AK117758; BAC42406.1; -
DR PIR; T06029; T06029.
KW Hypothetical protein.
SQ SEQUENCE 532 AA; 59135 MW; 0E2B76572C5D0DEC CRC64;

Alignment Scores:
Pred. No.: 0.385 Length: 532
Score: 115.00 Matches: 50
Percent Similarity: 41.36% Conservatve: 29
Best Local Similarity: 26.18% Mismatches: 29
Query Match: 7.01% Indels: 22
DB: 2 Gaps: 6

AF334735 (1-954) x Q9T069 (1-532)
QY 70 ACAACGGAAACCGGCGGACAGCTCGAGAGAAAGAGGTTCCATAGGCGAGTTCTTACC 129
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
226 ThrThrGluAsnGlySerAspGlyGluGlnGlnGluThrLysSerGluLeuAspSer 245
QY 130 AAGAAGATGCGATTCATCCATCTCCACACCCACTACCGAATTCACAA-----GGA 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246 LysThrGlyGlyGlyGlyGlySerAspSerAspGlyGluLeuLeuThrAsnLeuSer 265
QY 181 TTTTGGGAATCTTCTTGAAGGCTGACACCGAGATCTTGAGAGAGCAACCGACAAATATA 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
266 ThrSerAsnAlaThrGlu-----ThrThrGluSerSerGlySerAspGluSerGlySer 283
QY 241 CCAGCTTTTGACGACGCTTATTTTGAGCGCTTCTAGAGCCTTCTAGAGAAAGAGAGAAACCACTTT 300
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
284 SerGlyLysSerThrGlyThrGlnGlnThrLysAsnGluGluAspGluLysGluVal 303
QY 301 GATCCAGCAAGATGGGGAGTAGGTAGAGACCGCTTCTATAAC----- 345
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
304 GlnSerSerGluGluGluSerLysValLysGluSerGlyLysAsnGluLysAspAlaSer 323
QY 346 ----AATCATGCAATTCAGAGCAAGAACCACTCTGAGAAAGTGATCTTAAACAAGAGAG 402
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
324 SerSerGlnAspGluSerLysGluGluLysProGluAtgGlyLysLysGluSerSer 343
QY 403 TCTCATGATCTGGGAGAGAGAA-----GAGCATCATCTACC 441
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
344 SerGlnGlyGluGlyLysGluGluProGluLysArgGluLysGluAspSerSerSer 363

```

---

```

QY 442 ATCTAGACTCTTCTGAGGAAGATAAGAAAAAGAGAGGTTCGTCTGCTCAAAATCCAA 501
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
364 GlnGluGluSerLysGluGluProGluAsnLysGluLysGluLysGluLysSerSerGln 383
QY 502 GCTGCTTCCGGGACACATCCAGAGAGAGGAGGCAAGAAATGAAACAAATAGTCTT 561
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
384 Glu-----GluAsnGluLysGluThrGluLysGluLysGluLysGluLysSerSer 401
QY 562 CAA-----AATGAGAAAGAGGAAACCAAG 588
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
402 GlnGluGlyAsnGluAsnLysGluThrGluLys 412

RESULT 30
Q75SR9 PRELIMINARY; PRT; 992 AA.
AC Q75SR9;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Predicted protein.
GN Name=NCU05774.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamysseilis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Krystofova S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Osmani S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Piamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Eboole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABX01000353; EAA30798.1; -.
SQ SEQUENCE 992 AA; 111858 MW; 7443B6FDBDCDD182 CRC64;

Alignment Scores:
Pred. No.: 0.493 Length: 992
Score: 114.00 Matches: 37
Percent Similarity: 45.99% Conservatve: 26
Best Local Similarity: 27.01% Mismatches: 38
Query Match: 6.95% Indels: 36
DB: 2 Gaps: 6

AF334735 (1-954) x Q75SR9 (1-992)
QY 220 AGAGAGCAACCGGACCAATATACCACTTTTCGACGAGCCTATTTTGAGAGCCTCTAGAG 279
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
382 ArgGluAspAlaAspMetIleProGlu-----Ser 391
QY 280 AAAAGAGAGAAACCAACTTTGATCCAGCAAGATGGGGAGTAGAGTAGAGACCGCTTC 339
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
392 GluArgGluGluGluValAlaAspAlaGlyGln-----ThrAspValGluArgAsp 409
QY 340 TAT-----AACAAATCATGCTTCGAG 360
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
410 TyrGluGluArgAspGluMetLysSerSerGlySerSerSerSerSerSerGlyGluAsp 429
QY 361 GAGCAAGAACCACTCTGAGAAAAAGTATCTCTAAACAAGAGAGTCTCAGATATCTGGGAG 420

```







DR InterPro: IPR003439; ABC\_transporter.  
 DR Pfam: PF00005; ABC\_tran; 2.  
 DR ProDom: PD000006; ABC\_transporter; 2.  
 DR SMART: SM00382; AAA; 2.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER\_1; UNKNOWN\_2.  
 DR PROSITE: PS00893; ABC\_TRANSPORTER\_2; 2.  
 KW ATP-binding.  
 SQ SEQUENCE 845 AA; 95935 MW; C7889EF178522059 CRC64;

Alignment Scores:  
 Pred. No.: 0.723 Length: 845  
 Score: 112.00 Matches: 45  
 Percent Similarity: 41.72% Conservative: 23  
 Best Local Similarity: 27.61% Mismatches: 59  
 Query Match: 6.83% Indels: 36  
 DB: Gaps: 2

AF334735 (1-954) x Q8NE71 (1-845)

QY 130 AGAAGATGCGATTCATCTTCAACACCCACTACCGAATTCCA----- 174  
 D 102 LysLysLeuSerValProThrSerAspGluGluAspGluValProAlaProLysProArg 121  
 QY 175 -----CAAGATTTCGGAATCTCTTGAAGGCTGACACGCGAGATTCG 219  
 D 122 GlyGlyLysThrLysGly---GlyAsnValPheAlaAlaLeuIleGlnAspGlnSer 140  
 QY 220 AGAGACCAACCGACAAATATACCGCTTTTTCGACGAGCCTATTTTGAGAGCCTCTTAGAG 279  
 D 141 GluGluGlu-----Glu 144  
 QY 280 AAAAGAGAGAAACCAACTTTGATCCAGCAGAGATGGGGAGTAAGGTAGACACCGCTTC 339  
 D 145 GluGluGluLysHisProLysProLysProAla-----LysProGluLysAsnArg 160  
 QY 340 TATACAAATCGATTCATTCAGAGCAAGAACCACTGTGAGAAAGATGATCCTTAACAGAA 399  
 D 161 IleAsnLysAlaValProGluGluGlnGlnProAlaLeuLysGlyLysGlyLysGlu 180  
 QY 400 GAGTCTCAGATCTCTGGGAGGAGGAAGACATCAGTCACCATCTTAGACTCTTCTGAG 459  
 D 181 GluLysSerLysGlyLysAlaLysProGlnAsnLysPheAlaLeuAspAsnGluGlu 200  
 QY 460 GAAGATAAGGAAAGAGAGAGGTGTCTGTGTCACAAATCCAGTGTCTTCGCGGACAC 519  
 D 201 GluAspLysGlu---GluGluIleLysGluLys---GluProLysGlnGly--- 217  
 QY 520 ATAGCCAGAGAGGCAAGAAATGAACAAATAGTCTTCAAAATGAGGAAAGAG 579  
 D 218 -----LysGluLysAlaLysLysAlaGluGlnGlySerGluGluGlyGluGlyGlu 235  
 QY 580 GAAACCAAG 588  
 D 236 GluGluGlu 238

RESULT 34

QY Q7RRX0 PRELIMINARY; PRT; 3290 AA.  
 AC Q7RRX0; 26, Created)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Hypothetical protein.  
 GN Names: PY00597;  
 OS Plasmodium yoelii yoelii.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=73239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=17XNL;  
 RX PubMed=12368865; DOI=10.1038/nature01099;  
 RA Carlton J.M., Anguolui S.V., Suh B.B., Kooij T.W., Perteau M.,  
 Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,

RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
 RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,  
 RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,  
 RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,  
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,  
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,  
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
 RA Carucci D.J.;  
 RT "Genome sequence and comparative analysis of the model rodent malaria  
 parasite Plasmodium yoelii yoelii";  
 RL Nature 419:512-519 (2002).  
 CC -!- CAUTION: The sequence shown here is derived from an  
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 preliminary data.  
 CC EMBL: AABL01000162; EAA17150.1; -.  
 DR Hypothetical protein.  
 KW Hypothetical protein.  
 SQ SEQUENCE 3290 AA; 385404 MW; B078B911200C0666 CRC64;

Alignment Scores:

Pred. No.: 0.806 Length: 3290  
 Score: 112.00 Matches: 45  
 Percent Similarity: 42.05% Conservative: 37  
 Best Local Similarity: 23.08% Mismatches: 86  
 Query Match: 6.83% Indels: 27  
 DB: Gaps: 5

AF334735 (1-954) x Q7RRX0 (1-3290)

QY 187 AATCTTCTGAAGGCTGACACGCGAGATTCGAGAGACAA----- 228  
 D 2037 AsnLeuThrLysValIleAeArgAspIlePheAsnAsnGluPheGluValAsnIleAsp 2056  
 QY 229 -----CGGCAATATACCA----- 243  
 D 2057 IleIleLysAspGluIleProThrSerValHisThrGluProProLeuProThrProPro 2076  
 QY 244 ---GCTTTTGCAGCAGCCTATTTTGAGAGCCTCTTAGAGAAAGAGAGAAACCACTTT 300  
 D 2077 ThrSerThrThrValProIleGluIleGluLysGluGluLysGluIleGlu 2096  
 QY 301 GATCAGCAGAAATGGGGAGTAAGTAGAACCCCTCTTATACCAATCATGCTCGAG 360  
 D 2097 GluAspLysGluIleGluIleGluIleGluIleGluLysGluLysGluLysGluLys 2114  
 QY 361 GAGCAAGAACCACTGAGAAAGTATCTTAAAGAGAGAGTCTCAGATATCTGGAG 420  
 D 2115 GluGluLysGluIleGluLysGluLysGluLysGluLysGluLysGluLysGluLys 2133  
 QY 421 GAGGAAGAGACATCAGTCACCATCTTAGACTCTTCTGAGGAAGAGATAAGGAAAGAGAG 480  
 D 2134 GluGluGluLysGluValGluLysGluGluLysGluValGluLysGluGluLys 2153  
 QY 481 GTTGCTGTGTCAAAATCCAAAGCTGCTTCCGGGACACATAGCCAGAGAGAGCAAG 540  
 D 2154 GluValGluLysGluValGluLysGluValGluLysGluValGluLysGluValGluLys 2173  
 QY 541 AAAATGAAACCAATAGTCTTCAAAATGAGAAAGAGAGAGAGAGAGAGAGAGAGAG 600  
 D 2174 GluValGluLysGluGluGluLysGluGluLysGluGluLysGluGluLysGluGluLys 2193  
 QY 601 TTTTACCTCCAGAAACATGAAATAATCAATATCAATATCAATCAAC---CTTCTTATTAAGT 657  
 D 2193 sGluValGluLysGluGluLysGluLysSerAsnThrSerCysLysCysCysLeuValLys 2213  
 QY 658 CATTCTCTTCGAGAGAGAGATTTGATGTTGTGTAATAACA 700  
 D 2213 eileAsnAsnThrGlySerAsnGlyAsnCysIleSerValThr 2227  
 RESULT 35  
 QY Q7TU32 PRELIMINARY; PRT; 396 AA.  
 AC Q6TU32;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)

```

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE PKA type II regulatory subunit.
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Orthogastropoda;
OC Apogastropoda; Heterobranchia; Euthyneura; Opisthobranchia; Anaspidia;
OC Aplysioidae; Aplysiidae; Aplysia.
OX NCBI_TaxID=6500;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15014122; DOI=10.1523/JNEUROSCI.4331-03.2004;
RA Liu J., Hu J.Y., Schacher S., Schwartz J.H.;
RT "The two regulatory subunits of aplysia cAMP-dependent protein kinase
RT mediate distinct functions in producing synaptic plasticity.";
RL J. Neurosci. 24:2465-2474(2004).
DR EMBL; AY387673; AAR3637.1; -.
DR GO; GO:0005952; C:cAMP-dependent protein kinase complex; IEA.
DR GO; GO:0008603; F:cAMP-dependent protein kinase regulator act. .; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR002373; CAMP_kin
DR InterPro; IPR000595; cNMP_binding.
DR InterPro; IPR003117; RiIa.
DR Pfam; PF00027; cNMP_binding; 2.
DR Pfam; PF02197; RiIa; 1.
DR PRINTS; PR00103; CAMPKINASE.
DR SMART; SM00100; cNMP; 2.
DR SMART; SM00394; RiIa; 1.
DR PROSITE; PS00888; cNMP_BINDING_1; 2.
DR PROSITE; PS00889; cNMP_BINDING_2; 2.
DR PROSITE; PS00042; cNMP_BINDING_3; 2.
SQ SEQUENCE 396 AA, 44500 MW, 10C7E63A52A50D83 CRC64;

Alignment Scores:
Pred. No.: 0.752 Length: 396
Score: 111.50 Matches: 46
Percent Similarity: 43.33% Conservative: 32
Best Local Similarity: 25.56% Mismatches: 78
Query Match: 6.79% Indels: 24
DB: 2 Gaps: 6

AF334735 (1-954) x Q6TU32 (1-396)

QY 160 CACTACCGAATTCACAGAGTTTGGGAATCTTTGAGGCTGACACGCGAGATCTG 219
DB 2 AenpheGluileProProGlyLeuThrAspleuLeuGluGluPheThrValalavalleu 21

QY 220 AGAGAGCAACCGGCAATATACAGCTTTTGAGCAGCGCTATTTTGAGAGCGCTCTAGAG 279
DB 22 ArgThrArgProSerAsnLeuGluAlaPheAlaAlaAspTyrPheAsnAsnLeuAsnGlu 41

QY 280 AAAGAGAGAAACCAACTTTGATCCAGCAATGGGGAGTAAGGTAGAGACCGCTTC 339
DB 42 LysLysAsn-----GlyProLeuLysLeuSerGlyAlaLeu-----Argphe 55

QY 340 TATACAAATCATGATTCGAGGAGCAAGACACCATCTGAGAAAGTGATCTAAACAAGAA 399
DB 56 GlnGluAsnAspThrValasnIleleGluProSerAspGlyHisGlySerSerAsnGlu 75

QY 400 GAGTCTCAGATATCTGGGAAGCAAGACATCAGTCACCATCTTAGACTTCTCTGAG 459
DB 76 Asp-----AspAspAspSerPheValGlyProProSerSerTyr 89

QY 460 GAAGATAGAAAAGAGAGGTTGCTGCTGTCAAAATCCAAAGTCGCTCCGGGGACAC 519
DB 90 ArgAspArgArgLys---SerValSerAlaGluArgTyrAspProGlu-----104

QY 520 ATAGCCAGGAGCGCAAGAAATGAAACAAATAGTCTTCAAAATGAGGAAAGAG 579
DB 105 ---AlaAspAspGluGlyAspTyrThrLysValValasnProLysSerAspGlu-GlnAr 123

QY 580 GAAACAACATGAGGACACT-----GTTTATCTCTCCAGGAACATGAAGA 624
DB 123 GARGLeuAsnAspAlaIleLysHisIleLeuLeuPheArgSerLeuAspHisGluGl 143
QY 625 ATAATCCAAATCCATCAACCTTCTTATTATGTTCATTTCTCTGAGGAAGGAAGATT 682
DB 143 nmetGlnGluValleuAspAlaMetPheGluLysIleValasProGlyGluGluIle 162

RESULT 36
Q9LVA0 PRELIMINARY; PRT; 446 AA.
ID Q9LVA0
AC Q9LVA0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE DBJ|BAA90612.1 (Hypothetical protein At5g62390).
GN Name=At5g62390;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:31-63(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Tripp M., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB019235; BAA97203.1; -.
DR EMBL; AY093078; AAM13077.1; -.
DR EMBL; AY128766; AAM91166.1; -.
DR InterPro; IPR000048; IQ_region.
DR Pfam; PF00612; IQ; 1.
DR PROSITE; PS00096; IQ; 1.
KW Hypothetical protein.
SQ SEQUENCE 446 AA, 51567 MW, 95E3125F4CC67F8 CRC64;

Alignment Scores:
Pred. No.: 0.759 Length: 446
Score: 111.50 Matches: 35
Percent Similarity: 45.99% Conservative: 28
Best Local Similarity: 25.55% Mismatches: 59
Query Match: 6.79% Indels: 15
DB: 2 Gaps: 4

AF334735 (1-954) x Q9LVA0 (1-446)

QY 196 GAAGGCTGACACGCGAGATTCTGAGAGCAACCGGCAATATACCAAGCTTTTTCAGCA 255
DB 171 GluGlyLeuLysLeuAlaAlaLeuLysGluLysAlaLysAlaLysAlaIleAla 190

QY 256 GCCTATTTTGAGAGCTTCTGAGAGAAAGAGAGAAACCAACTTGTATCCAGCAGATGG 315
DB 191 AlaGluAlaGluLysLysLysAsnLysLysLysSerTyrAsn-----Trp 207

```

QY 316 GGGAGTAAAGTAAAGACCGCTTCTATAC-----AATCATGCAATC----- 357  
 DB 208 ThrThrGluValLysSerGluArgGluAlaGluValSerHisThrTyrIleIleLys 227  
 QY 358 -----GAGGACGACAGACCACTGTGAGAAAGTCAATCAAGAGAG 402  
 DB 228 AlaThrThrGlyGluLysLysLysHisGluGluLysGluLysLysLys 247  
 QY 403 TCTCAGATATCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 459  
 DB 248 GluThrLysSerLysLysLysLysLysLysLysLysLysLysLysLysLys 267  
 QY 460 GAAGATAGGAAAAAGAGAGGTTGCTGTGTCACAAATCAAGCTGCCTTCGGGGACAC 519  
 DB 268 GluAspAspGluSerSerGluHisGlyAlaIleValLysLysLysLysLysLys 287  
 QY 520 ATAGCCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 570  
 DB 288 AsnGlyAlaValArgThrLysLysGlyLysLysLysLysLysLysLysLys 304  
 RESULT 37  
 Q8K1M3  
 ID Q8K1M3 PRELIMINARY; PRT; 402 AA.  
 AC Q8K1M3;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE CAMP-dependent protein kinase type II-alpha regulatory chain (Protein  
 DE kinase, cAMP dependent regulatory, type II alpha).  
 GN Name=Prkar2a;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Testis;  
 RC Mowen K.A., Collard M.W.;  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RC MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Strausberg R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Richards S., Worley K.C., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickton M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RA Strausberg R.;  
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RA Director MGC Project;

RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF533977; AAM97688.1; -  
 DR EMBL; BC075623; AAH75623.1; -  
 DR EMBL; BC080276; AAH80276.1; -  
 DR HSSP; P12367; 1R2A.  
 DR MGD; MGI:108025; Prkar2a.  
 DR GO; GO:0005515; F:protein binding; IPI.  
 DR InterPro; IPR002373; CAMP\_kin.  
 DR InterPro; IPR000595; CNMP\_binding.  
 DR InterPro; IPR003117; Riia\_box\_N.  
 DR InterPro; IPR010916; TONB\_Box\_N.  
 DR Pfam; PF00027; CNMP\_binding; 2.  
 DR Pfam; PF02197; Riia; 1.  
 DR PRINTS; PR00103; CAMPKINASE.  
 DR SMART; SM00100; CNMP; 2.  
 DR SMART; SM00394; Riia; 1.  
 DR PROSITE; PS00888; CNMP\_BINDING\_1; 2.  
 DR PROSITE; PS00889; CNMP\_BINDING\_2; 2.  
 DR PROSITE; PS00042; CNMP\_BINDING\_3; 2.  
 DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; UNKNOWN\_1.  
 KW Kinase.  
 SQ SEQUENCE 402 AA; 45589 MW; 505FDD8B35F1BA00 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.831 Length: 402  
 Score: 111.00 Matches: 45  
 Percent Similarity: 38.86% Conservative: 30  
 Best Local Similarity: 23.32% Mismatches: 62  
 Query Match: 6.76% Indels: 56  
 DB: 2 Gaps: 7  
 AF334735 (1-954) x Q8K1M3 (1-402)  
 QY 157 ACCCACTACCGAATTCACAGGATTTGGGAATCTTTGAGGGCTGACACGCGAGATT 216  
 DB 2 SerHisIleGlnIleProAlaGlyLeuThrGluLeuLeuGlnGlyTyrThrValGluVal 21  
 QY 217 CTGAGAGAGCAACCGGACATATACAGCTTTTGCAGCAGCCTATTTTTCAGAGCTTTCTA 276  
 DB 22 LeuArgGlnGlnProProAspLeuValAspPheAlaValGluTyrPheThrArgLeuArg 41  
 QY 277 GAGAAAAGAGAGAAA-----ACCACTTT----- 300  
 DB 42 GluAlaArgArgGlnGlnSerAspThrPheIleValSerProThrThrPheHisThrGln 61  
 QY 301 -----GATCCAGCAGATGGGGAGTAAAGTAGTAGAC 333  
 DB 62 GluSerSerAlaValProValIleGluAspGlyGluSerAspSerAspSerGluAsp 81  
 QY 334 -----CGC 336  
 DB 82 AlaAspLeuGluValProValProSerLysPheThrArgArgValSerValCysAlaGlu 101  
 QY 337 TCTTATACAAATCATGATTCAGAGGAGCAAGACCACTGTGAGAAAGTAGATCCTTAAACAA 396  
 DB 102 ThrPheAsnProAspGluGluGluAspAspProArgValValHisProLysThr 121  
 QY 397 GAAGAGTCTCAGATATCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 456  
 DB 132 AspGluGlnArgCysArgLeuGlnGlnAlaCysLysAspIleLeuLeuPhe 138  
 QY 457 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 516  
 DB 139 ---LysAsnLeuAspGlnGlnLeuSerGlnVal---LeuAspAlaMetPheGlu--- 155  
 QY 517 CACATAGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 576  
 DB 156 -----LysIleValLysThrAspGluHisValIleAspGlnGly 168  
 QY 577 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 615  
 DB 169 AspAsp-----GlyAspAsnPheTyrValIleGlu 178

```

RESULT 38
O9CYV3 PRELIMINARY; PRT; 134 AA.
AC O9CYV3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 10, 11 days embryo whole body cDNA, RIKEN full-length
DE enriched library, Clone:2810440D03 product:10, 11 DAYS EMERYO CDNA,
DE RIKEN FULL-LENGTH ENRICHED LIBRARY, CLONE:2810440D03, FULL INSERT
DE SEQUENCE, full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44 (1999).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
RN [3]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
RN [4]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=2049374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630 (2000).
RN [5]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771 (2000).
RN [6]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,

```

```

RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK013274; BAB28762.1; -.
DR HSRF; P12367; IR2A.
DR GO; GO:0008603; F:camp-dependent protein kinase regulator act. .; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR003117; RIIA.
DR InterPro; IPR010916; TONB_Box_N.
DR Pfam; PF02197; RIIA; 1.
DR SMART; SM00394; RIIA; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN 1.
SQ SEQUENCE 134 AA; 15229 MW; 19C83760E6548A8 CRC64;

Alignment Scores:
Pred. No.: 0.841 Length: 134
Score: 110.50 Matches: 31
Percent Similarity: 42.74% Conservative: 19
Best Local Similarity: 26.50% Mismatches: 52
Query Match: 6.73% Indels: 15
DB: Gaps: 2

AF334735 (1-954) x O9CYV3 (1-134)
Oy 157 ACCCACTACCGAATTCCACAGGATTTGGGAATCTTCTTGAAGGGCTGACACGCGAGATT 216
Db 2 SerHisIleGlnIleProAlaGlyLeuThrGluLeuGlnGlyTyrThrValGluVal 21
Oy 217 CTGAGAGAGCAACCGGACAATATACAGCTTTTCAGAGCGCTATTTTGAGAGCGCTTCTTA 276
Db 22 LeuArgGlnGlnProProAspLeuValAspPheAlaValGluTyrPheThrArgLeuArg 41
Oy 277 GAGAAAGAGAGAGAAACCACTTTGATCCAGCAGATGGGGGAGTAAGGTAGAGACCGC 336
Db 42 GluAlaArgArgGlnGluSer-----AspThr 50
Oy 337 TTCTATATACAAAT-----CATGCATTCGAGGAGCAAGAACCACTTGAGAGAAAGT 384
Db 51 PheIleValSerProThrThrPheHisThrGlnGluSerSerAlaValProValIleGlu 70
Oy 385 GATCCTTAACCAAGAGAGTCTCAGATATCTGGGAGGAGGAGAGACATCAGTCACCATC 444
Db 71 GluAspGlyGluSerAspSerAspSerGluAspAlaAspLeuGluAspProIleAlaCys 90
Oy 445 TTAGACTCTTCTGAGGAGAGATAAGGAAAGAGAGGTTGCTGCTGTCAAA 495
Db 91 LeuLeuSerGluAspLysSerThrSerThrArgGlnLeuAlaProAsnLys 107

RESULT 39
O8QTD2 PRELIMINARY; PRT; 921 AA.
ID O8QTD2
AC O8QTD2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE WSSV315.
OS White spot syndrome virus (WSSV).
OC Viruses; dsDNA viruses, no RNA stage; Nimaviridae; Whispovirus.
OX NCBI_TaxID=92652;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=20517548; PubMed=11062040; DOI=10.1006/viro.2000.0597;
RA Tsai M.F., Yu H.T., Tzeng H.F., Leu J.H., Chou C.M., Huang C.J.,
RA Wang C.H., Lin J.Y., Kou G.H., Lo C.F.;
RT "Identification and characterization of a shrimp white spot syndrome
RT virus (WSSV) gene that encodes a novel chimeric polypeptide of
RT cellular-type thymidine kinase and thymidylate kinase.";
RL Virology 277:100-110 (2000).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=21844071; PubMed=11853398; DOI=10.1006/viro.2001.1273;
RA Chen L.L., Leu J.H., Huang C.J., Chou C.M., Chen S.M., Wang C.H.,
RA Lo C.F., Kou G.H.;
RT "Identification of a nucleocapsid protein (VP35) gene of shrimp white

```

RT spot syndrome virus and characterization of the motif important for RT targeting VP35 to the nuclei of transfected insect cells.";  
RL Virology 293:44-53 (2002).

[3] **SEQUENCE FROM N.A.**  
**RP Lo C.-P., Kou G.-H.;**  
**RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.**  
**RL EMBL; AF440570; AAL89183.1; --**  
**SR EMBL; AF440570; AAL89183.1; --**  
**DQ SEQUENCE 921 AA; 102933 MW; C483AE887175C345 CRC64;**

Q8VAWS	PRELIMINARY;	PRT;	922 AA.
ID	Q8VAWS		
AC	Q8VAWS		
DT	01-MAR-2002	(TrEMBLrel. 20, Created)	
DT	01-MAR-2002	(TrEMBLrel. 20, Last sequence update)	
DT	01-OCT-2002	(TrEMBLrel. 22, Last annotation update)	
DE	W5V260.		
OS	White spot syndrome virus (WSSV).		
OC	Viruses; dsDNA viruses, no RNA stage; Nimaviridae; Whispovirus.		
NCBI_TaxID=92652;			
[1]			
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21549311; PubMed=11689662;		
RX	DOI=10.1128/JVI.75.23.11811-11820.2001;		
RX	Yang F., He J., Lin X., Li Q., Pan D., Zhang X., Xu X.;		
RT	"Complete genome sequence of the shrimp white spot bacilliform virus."		
RT	J. Virol. 75:11811-11820(2001).		
RL	EMBL; AF332093; AAL33263.1; -		
DR	SEQUENCE 922 AA; 103062 MW;		B0A6200D890C3E64 CRC64;
SD			





## RESULT 43

```

Q6CCN1 PRELIMINARY; PRT; 312 AA.
AC Q6CCN1;
DT 25-OCT-2004 (TEMBLrel. 28, Created)
DT 25-OCT-2004 (TEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TEMBLrel. 28, Last annotation update)
DE Yarrowia lipolytica chromosome C of strain CLIB99 of Yarrowia
DE lipolytica.
GN ORFNames=YALI0C08052g;
OS Yarrowia lipolytica CLIB99.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=284591;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RG Genolures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boisrame A., Boyer J., Cattolico L., Confanioli E., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RA "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: CR82129; CAG81884.1; -.
DR InterPro: IPR006958; Mak16.
DR Pfam: PF04874; Mak16; 1.
SQ SEQUENCE 312 AA; 36198 MW; 56E5C30944201A6C CRC64;

Alignment Scores:
Pred. No.: 1-21 Length: 312
Score: 109.00 Matches: 46
Percent Similarity: 39.27% Conservative: 29
Best Local Similarity: 24.08% Mismatches: 72
Query Match: 6.64% Indels: 44
DB: 2 Gaps: 5

AF334735 (1-954) x Q6CCN1 (1-312)
QY 118 GCAGTTCTTACCAAGAGTGTGATTCATTCTCCACACCCCACTACCGAATTCACAA 177
Db 121 AlaValAlaGluArgileAlaMetLysGluGluAspThrHisTyr----- 136
QY 178 GGAATTTGGGAATCTTCTGAGCGGTGACACGGGAGATCTTGAGAGCAACCGGACAAT 237
Db 137 ---ValGlyValAlaProLysValLysArgGluAlaThrArgGluArg----- 152
QY 238 ATACCAAGCTTTTGACAGCGCTATTTTGAGAGCTTCTAGAGAAA----- 282
Db 153 ----LysAlaLeuAlaAlaLysValGluLysAlaIleGluLysGluLeuLeuAspArg 171
QY 283 -----AGAGAGAAAACCAACTTTGATCCAGCAGAAATGGGGAGCT 321
Db 172 LeuLysSerGlyAlaTyrGlyGluAsnProLeuAsnValAspGluLysValTrpLysLys 191
QY 322 AGGTAGAGACCGCTTCTATACAAATCATGCAATTCGAGAGCAAGAACCACTGAGAAA 381
Db 192 ValLeuGluGlyValGluGlyAlaAsnLysGluAspGluGluGluGluGlu 211

```

```

QY 382 AGTGATCTCTAAACAAGAGAGTCTCAGATATCTGGGAAGGAGGAGAGACATCAGTCACC 441
Db 212 GluAspGluGluGluGluGlyGlyGluValGluTyrValGluAspGluGluAspAsp 231
QY 442 ATCTTA----- 447
Db 232 AspLeuValGluLeuGluAspLeuGluLysTyrTrpLeuGlyLysSerAspGluSerGly 251
QY 448 -----GACTCTTCTGAGGAA-----GATAAGGAAAAAGAGAGGTTGCTGCTGTC 492
Db 252 AspGluThrAspAlaSerGluAspGluThrAspThrAspSerGluAspGluLysProLys 271
QY 493 AAAATCCAGCTGCTTCGCGGACACATAGCAGAGAGGAGGCAAGAAATGAAAAACA 552
Db 272 LysArgLysAlaAlaAspGlyLysProLysProLysThrLysLysArgLysGly 291
QY 553 AATAGTCTTCAAAATGAGGAAAAAGGAAAC 585
Db 292 ProSerValGluValGluTyrGluGluAspAsn 302

RESULT 44
KAPR_BLAEM
ID KAPR_BLAEM STANDARD; PRT; 403 AA.
AC P31320;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE CAMP-dependent protein kinase regulatory subunit (PKA regulatory
DE subunit).
GN Name=PKAR;
OS Blastocladiella emersonii (Aquatic fungus).
OC Eukaryota; Fungi; Chytridiomycota; Blastocladiaceae;
OC Blastocladiella.
OX NCBI_TaxID=4808;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92381035; PubMed=1512258;
RA Marques M.V., Gomes S.L.;
RT "Cloning and structural analysis of the gene for the regulatory
RT subunit of CAMP-dependent protein kinase in Blastocladiella
RT emersonii.";
RL J. Biol. Chem. 267:17201-17207(1992).
CC -! SUBUNIT: Tetramer, composed of 2 regulatory (R) and 2 catalytic
CC (C) subunits. In the presence of cAMP it dissociates into 2 active
CC monomeric C subunits and an R dimer that binds four cAMP
CC molecules.
CC -! DEVELOPMENTAL STAGE: Protein kinase activity is low in vegetative
CC cells, rising sharply during sporulation; during germination its
CC activity decreases to its original level.
CC -! SIMILARITY: Belongs to the CAMP-dependent kinase regulatory chain
CC family.
CC -! SIMILARITY: Contains 2 cyclic nucleotide-binding domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M81714; AAA33016.1; -.
CC EMBL: M81713; AAA33015.1; -.
CC PIR: A43435; A43435.
CC HSP: P12369; 1CX4.
CC InterPro: IPR002373; CAMP_kin.
CC InterPro: IPR00595; CNMP_binding.
CC InterPro: IPR003117; Riia.
CC Pfam: PF00027; CNMP_binding; 2.
CC Pfam: PF02197; Riia; 1.
CC PRINTS: PR00103; CAMPKINASE.
CC SMART: SM00100; CNMP; 2.

```





```

Alignment Scores:
Pred. No.: 1.68 Length: 475
Score: 107.50 Matches: 34
Percent Similarity: 44.17% Conservative: 19
Best Local Similarity: 28.33% Mismatches: 36
Query Match: 6.55% Indels: 31
DB: 2 Gaps: 4

AF334735 (1-954) x Q8W5 (1-475)
QY 169 ATTCCAGAAAGTTGGGAATCTTTGAAGGCTGACACGCGAGATTCTGAGAGAGCAA 228
D 10 ValProTyrGlyLeuLysThrLeuLeuGluGlyIleSerArgAlaValLeuLysThrAsn 29
QY 229 CCGGACATATACCGACTTTTCAGCAGCCTATTTTGAGAGCCTCTTAGAGAAAGAGAG 288
D 30 ProSerAsnIleAsnGlnPheAlaAlaAlaTyrPheGlnGluLeu----- 44
QY 289 AAAACCAACTTTGATCCAGCAAGTGGGGAGTAAAGGTAGAGACCGCTTCTATAACAAT 348
D 45 ---ThrMetTyrArgValGluLysTrpSer----- 53
QY 349 CATGCATTCGAGGACCAAGAACCCCTGAGAAAGTGATCTCTAAACAAGAA----- 399
D 54 -----GluGlyThrThrProGlnLysLysLeuGluCysLeuLysGluProGlyLys 70
QY 400 -----GAGTCTCAGATA-----TCTGGGAAGAGGAGAGAGACA 432
D 71 ThrSerValGluSerLysValProThrGlnMetGluLysSerThrAspThrAspGluAsp 90
QY 433 TCAGTCACCATCTTACGACTCTTCTGAGGAAGATAGGAAAGAGAGAGGTTCTGCTCTC 492
D 91 AsnValThrArgThrGluTyrSerAspLysThrThrGlnPheProSerValTyrAlaVal 110

RESULT 48
IF2_AQUAE
ID IF2_AQUAE STANDARD; PRT; 805 AA.
AC O67825;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Translation initiation factor IF-2.
GN Name=if2; OrderedLocusNames=AQ_2032;
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320; DOI=10.1038/32831;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aufay M., Huber R.,
RA Feldman R.A., Short J.M., Olsen G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus".
RL Nature 392:353-358(1998).
CC -!- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the IF-2 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; AE000769; AAC07794.1; -.
DR PIR; E70474; E70474.
DR HSSP; O26359; 1G7S.
DR HAMAP; MF_00100; -.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000178; IF2.
DR InterPro; IPR006847; IF2_N.
DR InterPro; IPR000795; ProSyn_GTPbind.
DR InterPro; IPR005225; Small_GTP.
DR InterPro; IPR009000; Translat_factor.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 2.
DR Pfam; PF04760; IF2_N; 2.
DR PRINTS; PR00315; ELONGATNFTCT.
DR ProDom; PD18100; IF2; 1.
DR TIGRFAMs; TIGR00487; IF-2; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS01176; IF2; 1.
KW Complete proteome; GTP-binding; Initiation factor;
KW protein biosynthesis.
FT DOMAIN 309 458 G-domain.
FT NP_BIND 315 322 GTP (By similarity).
FT NP_BIND 362 366 GTP (By similarity).
FT NP_BIND 416 419 GTP (By similarity).
SQ SEQUENCE 805 AA; 91600 MW; 8A1BF300C69CA51E CRC64;

Alignment Scores:
Pred. No.: 1.76 Length: 805
Score: 107.50 Matches: 53
Percent Similarity: 42.49% Conservative: 29
Best Local Similarity: 27.46% Mismatches: 74
Query Match: 6.55% Indels: 37
DB: 1 Gaps: 8

AF334735 (1-954) x IF2_AQUAE (1-805)
QY 73 ACCGAAACGGCGGCACCGAGCTCGAGAGAAAGAGGAGTCTCATAGGAGTCTTACCAAG 132
D 4 ThrLysThrLysArgValSerValAlaLysGluLeu-----GlyValLysSerLys 21
QY 133 AAGATGTGATTCATTCCTCCACACCCACTACCGAATTCACCAAGGA----- 180
D 22 GluIle---IleGluPheLeuAsnGluTyrTyrProArgProAspGlyLysProTyrLys 40
QY 181 -----TTGGGAATCTT 192
D 41 AlaSerHisGlyLeuAspGluGlnAlaLeuGluMetIleTyrAspAlaPheGlyIleLys 60
QY 193 CTTGAGGGCTGACACCGGAGATTCTGAGAGAGCAACCGGACATATACCGAGCTTTTGCA 252
D 61 GluGluGluLysGluGluValValThrGluGln-----AlaGlnAla 75
QY 253 GCAGCCTATTTTGAGAGCCTTTCTAGAGAAAGAGAGAAACCACTTTGATCCAGCAGAA 312
D 76 ProAlaGluValGlu-----GluLysLysGluGluLysLysGluGluValIle 92
QY 313 TGGGGGAGTAGGTAGGAGACCGCTTCTATAACAATCATGCATTCGAGGAGCAAGAACCA 372
D 93 ValGluGluValValGluGluLys---LysProGluValIleValGluGluIleGluGlu 111
QY 373 CCTGAGAAAGTGTATCTCTAACAGAGAGCTCTCAGATATCTCGGAAGGAGGAGAGACA 432
D 112 LysLysGluGluGluLysLysGluLysProLysLysSerValGluGluLeu 131
QY 433 TCAGTCACTCTTAGACTCTTCTGAGAAAGTAAAGAAAAAGAGAGTTGCTGCTGTC 492
D 132 IleLysGluIleLeuGluLysLysGluLysGluLysLysLysValGluLysGlu 151
QY 493 AAATCCAGCTGCTTCCCG---GGACACATGCCAGAGAGAGGAGGCAAGAAATGAA 549
D 152 ArgLysGluLysValValGluValValGluValLysLysGluLysGluLys 171
QY 550 ACAAATAGTCTTCAAAATGAGAAAGAGGAAACCAAG 588

```

```
Db 172 -----LyseGluLysLysGluGluGluLys 180
RESULT 49
Q7RK82
ID Q7RK82 PRELIMINARY; PRT; 881 AA.
AC Q7RK82;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Maeb1 (Fragment).
GN Name=PY03020;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=1236865; DOI=10.1038/nature01099;
RA Carlton J.M., Anguoli S.V., Suh B.B., Kooij T.W., Perteu M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallow S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RA "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; ABL01000854; EAA22546.1; -.
DR HSSP; P17676; IHJB.
FT NON TER 881
SQ SEQUENCE 881 AA; 105238 MW; F6B81ADD18C98B79 CRC64;

Alignment Scores:
Pred. No.: 1.77 Length: 881
Score: 107.50 Matches: 50
Percent Similarity: 46.88% Conservative: 40
Best Local Similarity: 26.04% Mismatches: 82
Query Match: 6.55% Indels: 20
DB: 2 Gaps: 9

AF334735 (1-954) x Q7RK82 (1-881)
QY 39 AAGAAAAACGGTTACCCAGCACTAGAAAAACAAACCGGCGGCGCAGCAGCTCGGA 98
Db 285 LysLysLysGluGluAspValAspPheAlaAsnLeuArgAsnLysLysGluAsp 304
QY 99 GAGAAAGAGGTTCCATAGGCGATCTTACCAAGAGATGTCGATTCATTCCTCAACAC 158
Db 305 ValAspPheAlaAsnLeuArgAsnLysLysGluGlu-AspValAspPheSerAsnLe 324
QY 159 CCNACTACCAANTCCACAA-----CGATTGGGAATCTCTTGAAGGCGTGACA-- 207
Db 324 uArgAsnLysLysLysGluGluAspValAspPheGlyAsnLeuArgAsnLysLysAsnAs 344
QY 208 -CGCAGAGATCTGAGAGAGCAACCGGACAAATATACCGATTTTGCAGCAGCCTATTTCGA 266
Db 344 pThrGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 364
QY 267 GAGCCTCTTAGAGAAAGAGAGAAACCAACTTTGATCCAGAGATGGGGGAGTAAGGT 326
Db 364 nLysGlnLeu---LysArgGlnGluThrGluLysSerGln-----AsnGlyLysLys 380
QY 327 AGAAGACCGCTTCTATACCAATCATTCATTCGAGGAGGAGCAAGAACCACTGAG-----AA 380
Db 380 uGluGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 400

RESULT 50
Q8I590
ID Q8I590 PRELIMINARY; PRT; 3193 AA.
AC Q8I590;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PFL1600C;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2225705; PubMed=1236864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.S., Nene V., Shallow S.J., Suh B., Peterson J., Anguoli S.,
RA Perteu M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RA "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Hyman R.W., Fung E., Conway A., Kurdi O., Mao J., Miranda M.,
RA Nakao B., Rowley D., Tamaki T., Wang F., Davis R.W.;
RA Submitted (JAN-2003) to the EMBL/GenBank/DDJ databases.
DR EMBL; AB014849; AAN36406.1; -.
KW Hypothetical protein.
SQ SEQUENCE 3193 AA; 377925 MW; 46E6B21F921C5307 CRC64;

Alignment Scores:
Pred. No.: 1.96 Length: 3193
Score: 107.50 Matches: 51
Percent Similarity: 39.52% Conservative: 32
Best Local Similarity: 24.22% Mismatches: 63
Query Match: 6.55% Indels: 64
DB: 2 Gaps: 10

AF334735 (1-954) x Q8I590 (1-3193)
QY 39 AAGAAAAACGGTTACCCAGCACTAGAAAAACAAACCGGCGGCGCAGCAGCTCGGA 98
Db 1641 LysLysAsnAspValThrAspAsnAsnGluAsnAsnGlyLysArgAsnLysLysLeuArg 1660
QY 99 -----GAGAAAGAGGTTCCATA-----GCG 119
Db 1661 LysAlaLysThrPheGlyThrValLysLysGlyThrLysLysLysLysLysLysLys 1680
QY 120 AGTTCTTACCAAGAGATGTCGATTCCTTC-----AACAC 158
```

```
Db 1681 GluSer-GlnArgIleLysArgThrProPheCysIleTyrSerLeuAsnAspIleCysAs 1700
QY 159 CCACCTACCGAATTCACAGGATTTGGGAATCTCTTGAAGGCTGACACGC----- 210
Db 1700 nHisLeuArgTyrGlyLysSerAlaAsnSerIleAsnGlyMetIleLysThrSerTh 1720
QY 211 -----GAGATTCTGAGAGACCAACCGGACAATATACCAGCTTTTCAGCAGCCTA 260
Db 1720 rCysAsnCysGluIle-----HisAsnAspMetLysPheValLysGlnTy 1735
QY 261 TTTTGAGAGCCTTCAGAGAAAGAGAGAAACCAACTTTGATCCAGCAGAAATGGGGAG 320
Db 1735 rIleAsnPheLeuIleAlaGlnLysGluLysGluArgIleAsn----- 1749
QY 321 TAAGGTAGAGACCCCTTCTATAACAATCATGCTTCGAGGAGCAAGAACCACTGAGAA 380
Db 1750 -----LysGluLysLysIleAsn-----GluLysGluLysAsnAspG1 1763
QY 381 AAGTCATCTAAACAAGAGAGTCTCAGATATCTGGGAAGGAGGAGAGACATCAGTCAC 440
Db 1763 uGlyAsnLysLysGluGluSerGln-----LysGluGluGluSer----- 1777
QY 441 CATCTTAGACTTCTGTAGGAGATPAGGAAAGAGAGGTTGCTGCTGTCAAAATCCA 500
Db 1778 -----GlnLysGluGluSerGlnLysGluGluSerLys----- 1790
QY 501 AGCTGCCTTCGGGGACACATAGCCAGAGAGGAGCAAGAAATGAAACAAATAGTCT 560
Db 1791 -----AsnGluGluGluSerLysAsnGluGluAsnAsnIleLy 1803
QY 561 TCAAAATGAGAAAGAGGAAACAAAG 588
Db 1803 sAspAsnGluGlnLysArgLysAsnLys 1812
```

Search completed: February 15, 2005, 10:06:29  
Job time : 189.5 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 15, 2005, 02:05:31 ; Search time 201 Seconds  
(without alignments)  
4860,931 Million cell updates/sec

Title: AF334735

Perfect score: 307

Sequence: 1 TCGCCCTTCGCGCGCG.....AAAAAAAAAAAAAAAAAAAA 954

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Word size: 1

Total number of hits satisfying chosen parameters: 3224408

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool\_p/AF334735/runat\_14022005\_160237\_19339/app\_query.fasta\_1.1095  
-DB=Uniprot\_03 -Qfmt=fastan -SUFFIX=oligo.rup -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human4g/cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=AF334735 @CGN 1 1 244 @runat\_14022005\_160237\_19339 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Uniprot 03.\*

1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	151	49.2	151	1 SP17 HUMAN	Q15506 homo sapien
2	52	30.0	151	1 SP17 MACFA	Q19021 macaca fasc
3	91	29.6	163	1 SP17 PAPHA	Q95230 papio hamad
4	71	23.1	143	2 Q9BDQ6	Q9BDQ6 macaca mula
5	44	14.3	146	1 SP17 RABIT	P36425 oryctolagus
6	44	14.3	148	2 Q9ZIK2	Q9ZIK2 rattus norv
7	44	14.3	149	1 SP17 MOUSE	Q62252 mus musculu
8	44	14.3	153	1 SP17 MACEU	Q62770 macropus eu
9	44	14.3	179	1 SP17 MONDO	Q62771 monodelphis
10	40	13.0	147	2 Q9TU08	Q9TU08 ovis aries
11	33	10.7	212	2 Q9XT29	Q9XT29 callithrix
12	15	4.9	141	2 Q6SJ92	Q6SJ92 ornithorhyn
13	13	4.2	495	2 Q6DIJ9	Q6DIJ9 xenopus tro
14	11	3.6	1012	2 Q7SH99	Q7SH99 neurospora
15	10	3.3	1148	2 Q8FN75	Q8FN75 corynebacte
16	10	3.3	1448	2 Q9XZE9	Q9XZE9 hydra atten

17	9	2.9	139	2	Q7QHMI	Q7qhml anopheles g
18	9	2.9	145	1	STIA_XENLA	Q09006 xenopus lae
19	9	2.9	471	2	Q73MIS	Q73mi5 treponema d
20	9	2.9	549	2	Q9G910	Q9g910 ochromonas
21	9	2.9	863	2	Q9BTP1	Q9bcp1 homo sapien
22	9	2.9	1040	2	Q8NH2	Q8nhn2 homo sapien
23	9	2.9	6620	2	Q96AA2	Q96aa2 homo sapien
24	8	2.6	48	2	Q7M2U5	Q7m2u5 bubalus bub
25	8	2.6	48	2	Q8F422	Q8f422 leptospira
26	8	2.6	66	2	Q49241	Q49241 mycoplasma
27	8	2.6	67	2	Q72W40	Q72w40 leptospira
28	8	2.6	67	2	Q8F9S8	Q8f9s8 leptospira
29	8	2.6	88	2	Q61WB2	Q61wb2 arabidopsis
30	8	2.6	102	2	Q8C5Y8	Q8c5y8 mus musculu
31	8	2.6	105	2	Q35129	Q35129 neurospora
32	8	2.6	110	2	Q7QP28	Q7qp28 giardia lam
33	8	2.6	112	2	Q8BYC1	Q8byc1 mus musculu
34	8	2.6	112	2	Q9CUV5	Q9cuv5 mus musculu
35	8	2.6	123	2	Q653R8	Q653r8 oryza sativ
36	8	2.6	133	2	Q6PIM9	Q6pim9 homo sapien
37	8	2.6	139	1	YED3_YEAST	P32633 saccharomyc
38	8	2.6	140	2	Q8BRM4	Q8brm4 mus musculu
39	8	2.6	143	2	Q9C4A7	Q9c4a7 tricholoma
40	8	2.6	150	2	Q8ENS8	Q8ens8 oceanobacill
41	8	2.6	151	2	Q9AQY4	Q9aqy4 guillardia
42	8	2.6	158	2	Q88B26	Q88b26 pseudomonas
43	8	2.6	166	2	Q9BGP4	Q9bgp4 macaca fasc
44	8	2.6	170	2	Q9NXR4	Q9nxr4 homo sapien
45	8	2.6	177	2	Q8L3Q3	Q8l3q3 oryza sativ

#### ALIGNMENTS

##### RESULT 1

SP17\_HUMAN  
ID SP17\_HUMAN STANDARD; PRT; 151 AA.  
AC Q15506; Q9BXF7;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 25-OCT-2004 (Rel. 45, Last annotation update)  
DE Sperm surface protein Sp17 (Sperm autoantigenic protein 17) (Sperm protein 17) (Sp17-1).  
GN Name=SPAL7; Synonyms=SP17;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=96305346; PubMed=8688458; DOI=10.1016/0167-4781(96)00077-2;  
RA Lea I.A., Richardson R.T., Widgren E.E., O'Rand M.G.;  
RT "Cloning and sequencing of cDNAs encoding the human sperm protein, Sp17.";  
RL Biochim. Biophys. Acta 1307:263-266(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=22280349; PubMed=12393185; DOI=10.1016/S0167-4781(02)00478-5;  
RA Buchli R., De Jong A., Robbins D.L.;  
RT "Genomic organization of an intron-containing sperm protein 17 gene (Sp17-1) and an intronless pseudogene (Sp17-2) in humans: a new model.";  
RL Biochim. Biophys. Acta 1578:29-42(2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=23388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler N.K., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore S.I., Wang J., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,



Alignment Scores:  
 Pred. No.: 1.23e-85 Length: 151  
 Score: 92.00 Matches: 92  
 Percent Similarity: 100.00% Conservations: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 29.97% Indels: 0  
 DB: 1 Gaps: 0

AF334735 (1-954) x SP17\_MACFA (1-151)

QY 136 ATGTCGATTCATCTCCACACCCACTACCGAATTCACAGATTGGGATCTCTT 195  
 DB 1 MetSerileProPheSerAsnThrHisTyArgileProGlnGlyPheGlyAsnLeu 20  
 QY 196 GAAGGGCTGACCGCAGATTCTGAGAGACCAACCGACAAATATACAGCTTTTCGAGCA 255  
 DB 21 GluGlyLeuThrArgGluLeuLeuArgGluGlnProAsnAsnleProAlaPheAla 40  
 QY 256 GCCTATTTTTCGAGCCCTTCTAGAGAAAGAGAGAAACCAACTTTGATCCAGCAGATGG 315  
 DB 41 AlaTyPheGluSerLeuLeuGluLysArgGluLysThrAsnPheAspProAlaGluTrp 60  
 QY 316 GCGAGTAAGTAGAGACCCCTTCTATACAAATCATGCTTCGAGAGCAGAACCACTT 375  
 DB 61 GlySerLysValGluAspArgPheTyAsnAsnHisAlaPheGluGlnGluProPro 80  
 QY 376 GAGAAAGTGCATCTTAAACAGAGAGAGTCTCAGATA 411  
 DB 81 GluLysSerAspProLysGlnGluSerGlnle 92

RESULT 3

SP17\_PAPHA STANDARD; PRT; 163 AA.

AC Q95230;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Sperm surface protein Sp17 (Sperm autoantigenic protein 17).  
 GN Names=SP17; Synonyms=SP17;  
 OS Papio hamadryas (Hamadryas baboon).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Papio.  
 OX NCBI\_TaxID=9557;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=97264490; PubMed=9110316;  
 RA DOI=10.1002/(SICI)1098-2795(199705)47:1<66::AID-MRD9>3.0.CO;2-O;  
 RA Adoyo P.A., Lea I.A., Richardson R.T., Widgren E.E., O'Rand M.G.;  
 RT "Sequence and characterization of the sperm protein Sp17 from the  
 RT baboon.";  
 RL Mol. Reprod. Dev. 47:66-71(1997).  
 CC -!- FUNCTION: Sperm surface zona pellucida binding protein. Helps to  
 CC bind spermatozoa to the zona pellucida with high affinity. Might  
 CC function in binding zona pellucida and carbohydrates (By  
 CC similarity).  
 CC -!- SUBUNIT: Homodimer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Membrane-associated (Potential).  
 CC -!- TISSUE SPECIFICITY: Testis- and sperm-specific.  
 CC -!- SIMILARITY: Contains 1 IQ domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U75209; BAB38534.1; -;  
 CC HSSP; P12367; IR2A.

DR InterPro; IPR000048; IQ region.  
 DR InterPro; IPR003117; RIIA.  
 DR Pfam; PF0612; IQ; 1.  
 DR SMART; SM00197; RIIA; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00394; RIIA; 1.  
 DR PROSITE; PS00096; IQ; 1.  
 KW Membrane.  
 FT DOMAIN 114 143 IQ.  
 SQ SEQUENCE 163 AA; 18755 MW; A890B59B3F6469BD CRC64;

Alignment Scores:  
 Pred. No.: 1.33e-84 Length: 163  
 Score: 91.00 Matches: 91  
 Percent Similarity: 100.00% Conservations: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 29.64% Indels: 0  
 DB: 1 Gaps: 0

AF334735 (1-954) x SP17\_PAPHA (1-163)

QY 136 ATGTCGATTCATCTCCACACCCACTACCGAATTCACAGATTGGGATCTCTT 195  
 DB 1 MetSerileProPheSerAsnThrHisTyArgileProGlnGlyPheGlyAsnLeu 20  
 QY 196 GAAGGGCTGACCGCAGATTCTGAGAGACCAACCGACAAATATACAGCTTTTCGAGCA 255  
 DB 21 GluGlyLeuThrArgGluLeuLeuArgGluGlnProAsnAsnleProAlaPheAla 40  
 QY 256 GCCTATTTTTCGAGCCCTTCTAGAGAAAGAGAGAAACCAACTTTGATCCAGCAGATGG 315  
 DB 41 AlaTyPheGluSerLeuLeuGluLysArgGluLysThrAsnPheAspProAlaGluTrp 60  
 QY 316 GCGAGTAAGTAGAGACCCCTTCTATACAAATCATGCTTCGAGAGCAGAACCACTT 375  
 DB 61 GlySerLysValGluAspArgPheTyAsnAsnHisAlaPheGluGlnGluProPro 80  
 QY 376 GAGAAAGTGCATCTTAAACAGAGAGAGTCTCAG 408  
 DB 81 GluLysSerAspProLysGlnGluSerGln 91

RESULT 4

Q9BDQ6 PRELIMINARY; PRT; 143 AA.

AC Q9BDQ6;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Sperm protein 17 (Fragment).  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Macaca.  
 OX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Spleen;  
 RX MEDLINE=22280349; PubMed=12393185; DOI=10.1016/S0167-4781(02)00478-5;  
 RA Buchli R., De Jong A., Robbins D.L.;  
 RT "Genomic organization of an intron-containing sperm protein 17 gene  
 RT (Sp17-1) and an intronless pseudogene (Sp17-2) in humans: a new  
 RT model.";  
 RL Biochim. Biophys. Acta 1578:29-42(2002).  
 DR EMBL; AF334809; AAK28124.1; -;  
 DR HSSP; P12367; IR2A.  
 DR GO; GO:0008603; F: cAMP-dependent protein kinase regulator act. . . ; IEA.  
 DR GO; GO:0007165; P: signal transduction; IEA.  
 DR InterPro; IPR000048; IQ region.  
 DR InterPro; IPR003117; RIIA.  
 DR Pfam; PF0612; IQ; 1.  
 DR Pfam; PF02197; RIIA; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00394; RIIA; 1.

DR PROSITE; PS50096; IQ; 1.  
FT NON\_TER 1 143  
FT NON\_TER 143 143  
SQ SEQUENCE 143 AA; 16424 MW; 70B538F7F876B465 CRC64;

## Alignment Scores:

Pred. No.: 7,48e-64 Length: 143  
Score: 71.00 Matches: 71  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 23.13% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x Q9BDQ6 (1-143)

QY 154 AACACCCATACCGAATTCACAGGATTGGGAATCTCTTGAAGGGCTGACACGGAG 213  
DB 1 AsnThrHisTyrArgIleProGlnGlyPheGlyAsnLeuGluGlyLeuThrArgGlu 20  
QY 214 ATTCGAGAGACCAACCGACAAATATACAGCTTTTGCAGCAGCCTATTTTGAGAGCCTT 273  
DB 21 IleuUargGluGlnProAspAsnIleProAlaPheAlaAlaAlaTyrPheGluSerLeu 40  
QY 274 CTAGAGAAAGAGAGAAAACCACTTTGATCCAGCAGAAATGGGGAGTAGAGGTAGAAGAC 333  
DB 41 LeuGluYsArgGluYsThrAsnPheAspProAlaGluTrpGlySerIysValGluAsp 60  
QY 334 CGTCTCTAATCAATCATGCTTCGAGAGACAA 366  
DB 61 ArgPheTyrAsnAsnHisAlaPheGluGluGln 71

## RESULT 5

SP17\_RABIT  
ID SP17\_RABIT STANDARD; PRT; 146 AA.

AC F36425;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Sperm surface protein Sp17 (Sperm autoantigenic protein 17).  
GN Names=SPA17; Synonyms=Sp17;  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 113-126.  
RC STRAIN=New Zealand white; TISSUE=Testis;  
RX MEDLINE=95046885; PubMed=7525387; DOI=10.1006/dbio.1994.1285;  
RA Richardson R.T., Yamazaki N., O'Rand M.G.;  
RT "Sequence of a rabbit sperm zona pellucida binding protein and  
RT localization during the acrosome reaction."  
RL Dev. Biol. 165:688-701(1994).  
CC -!- FUNCTION: Sperm surface zona pellucida binding protein. Helps to  
CC bind spermatozoa to the zona pellucida with high affinity. Might  
CC function in binding zona pellucida and carbohydrates.  
CC -!- SUBUNIT: Homodimer.  
CC -!- SUBCELLULAR LOCATION: Membrane-associated (Potential).  
CC -!- TISSUE SPECIFICITY: testis- and sperm-specific.  
CC -!- PTM: The N-terminus is blocked.  
CC -!- SIMILARITY: Contains 1 IQ domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z20655; CAA79674.1; -.  
DR PIR; I46506; I46506.  
DR HSSP; P12367; 1R2A.  
DR InterPro; IPR000048; IQ\_region.

DR InterPro; IPR003117; RIIa.  
DR Pfam; PF00612; IQ; 1.  
DR Pfam; PF02197; RIIa; 1.  
DR SMART; SM00015; IQ; 1.  
DR SMART; SM00394; RIIa; 1.  
DR PROSITE; PS50096; IQ; 1.  
KW Direct protein sequencing; Membrane.  
FT DOMAIN 110 139  
SQ SEQUENCE 146 AA; 16891 MW; 3FF302B883D9566F CRC64;

## Alignment Scores:

Pred. No.: 7,48e-36 Length: 146  
Score: 44.00 Matches: 44  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 14.33% Indels: 0  
DB: 1 Gaps: 0

AF334735 (1-954) x SP17\_RABIT (1-146)

QY 136 ATGTCGATTCATTCCTCCACACCCCTACCGAATTCACAGGATTGGGAATCTTCTT 195  
DB 1 MetSerIleProPheSerAsnThrHisTyrArgIleProGlnGlyPheGlyAsnLeu 20  
QY 196 GAAGGGCTGACACGGCAGATTCTGAGAGAGCAACCGCAATATACAGCTTTTCGAGCA 255  
DB 21 GluGlyLeuThrArgGluLeuLeuArgGluGlnProAspAsnIleProAlaPheAla 40  
QY 256 GCCTATTTTGAG 267  
DB 41 AlaTyrPheGlu 44

## RESULT 6

Q9Z1K2  
ID Q9Z1K2 PRELIMINARY; PRT; 148 AA.

AC Q9Z1K2;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Sp17 protein.  
GN Name=sp17;  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Frayne J., Jury J.A., Hall L.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ131888; CAA10524.1; -.  
DR HSSP; P12367; 1R2A.  
DR GO; GO:0008603; F:cAMP-dependent protein kinase regulator act. . ; IEA.  
DR GO; GO:0007165; P:signal transduction; IEA.  
DR InterPro; IPR000048; IQ\_region.  
DR InterPro; IPR003117; RIIa.  
DR Pfam; PF00612; IQ; 1.  
DR Pfam; PF02197; RIIa; 1.  
DR SMART; SM00015; IQ; 1.  
DR SMART; SM00394; RIIa; 1.  
DR PROSITE; PS50096; IQ; 1.  
SQ SEQUENCE 148 AA; 17096 MW; 5E0BC9D4C86PD585 CRC64;

## Alignment Scores:

Pred. No.: 7,47e-36 Length: 148  
Score: 44.00 Matches: 44  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 14.33% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x Q9Z1K2 (1-148)

```

QY 136 ATGTGAGTCCATTCCTCCAAACACCCACTACCGAATTCACCAAGGATTTGGGAATCTTCTT 195
Db 1 MetSerileProPheSerAenThrHisTyArgileProGlnGlyPheGlyAenLeuLeu 20
QY 196 GAAGGGCTGACCGCGAGATTCGAGAGACGACACCGACCAATATACACGCTTTTGCAGCA 255
Db 21 GluGlyLeuThrArgGluIleLeuArgGluInProAspAsnIleProAlaPheAlaAla 40
QY 256 GCCTATTATTGAG 267
Db 41 AlaTyPheGlu 44

RESULT 7
ID SP17_MOUSE STANDARD; PRT; 149 AA.
AC Q62252;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE Sperm surface protein Sp17 (Sperm autoantigenic protein 17).
GN Name=Sp17; Synonyms=Sp17;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

SEQUENCE FROM N.A.
[1]
SPRAIN=BALB/c; TISSUE=Testis;
RX MEDLINE=9603129; PubMed=7578682;
RA Kong M., Richardson R.T., Widgren E.B., O'Rand M.G.;
RT "Sequence and localization of the mouse sperm autoantigenic protein,
RL Sp17."
RA Biol. Reprod. 53:579-590(1995).
[2]
SEQUENCE FROM N.A.
[1]
TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Sperm surface zona pellucida binding protein. Helps to
CC bind spermatozoa to the zona pellucida with high affinity. Might
CC function in binding zona pellucida and carbohydrates (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Membrane-associated (Potential).
CC -!- TISSUE SPECIFICITY: Testis- and sperm-specific.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```

```

CC EMBL; Z46299; CA86455.1; -.
DR EMBL; BC059727; AAHS9727.1; -.
PIR; S49527; S49527.
DR HSRP; P12367; IR2A.
DR MGD; MGI:1333778; Spal17.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR003117; RIIA.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF02197; RIIA; 1.
DR PROSITE; PS50096; IQ; 1.
KW Membrane.
FT DOMAIN 112 141 IQ.
SQ SEQUENCE 149 AA; 17296 MW; C7E05D111D6AF0DC CRC64;

Alignment Scores:
Pred. No.: 7,46e-36 Length: 149
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.33% Indels: 0
DB: 1 Gaps: 0

AF334735 (1-954) x SP17_MOUSE (1-149)
QY 136 ATGTGAGTCCATTCCTCCAAACACCCACTACCGAATTCACCAAGGATTTGGGAATCTTCTT 195
Db 1 MetSerileProPheSerAenThrHisTyArgileProGlnGlyPheGlyAenLeuLeu 20
QY 196 GAAGGGCTGACCGCGAGATTCGAGAGACGACACCGACCAATATACACGCTTTTGCAGCA 255
Db 21 GluGlyLeuThrArgGluIleLeuArgGluInProAspAsnIleProAlaPheAlaAla 40
QY 256 GCCTATTATTGAG 267
Db 41 AlaTyPheGlu 44

RESULT 8
SP17_MOUSE STANDARD; PRT; 153 AA.
ID SP17_MACEU STANDARD; PRT; 153 AA.
AC Q62770;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Sperm surface protein Sp17 (Sperm autoantigenic protein 17).
GN Name=SP17; Synonyms=SP17;
OS Macropus eugenii (Tamar wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
OX NCBI_TaxID=9315;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Wen Y., O'Rand M.G.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Sperm surface zona pellucida binding protein. Helps to
CC bind spermatozoa to the zona pellucida with high affinity. Might
CC function in binding zona pellucida and carbohydrates (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Membrane-associated (Potential).
CC -!- TISSUE SPECIFICITY: Testis- and sperm-specific (By similarity).
CC -!- SIMILARITY: Contains 1 IQ domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF054289; AAC08024.1; -.

```

DR HSSP; P12367; 1R2A.  
 DR InterPro; IPR000048; IQ region.  
 DR InterPro; IPR003117; RIIa.  
 DR Pfam; PF00612; IQ; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00394; RIIa; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00394; RIIa; 1.  
 DR PROSITE; PSS0096; IQ; 1.  
 KW Membrane.  
 FT DOMAIN 122 151 IQ.  
 SQ SEQUENCE 153 AA; 17327 MW; AD3A7C6B4D4E1B65 CRC64;  
 Alignment Scores:  
 Pred. No.: 7 43e-36 Length: 153  
 Score: 44.00 Matches: 44  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 14.33% Indels: 0  
 DB: 1 Gaps: 0

AF334735 (1-954) x SP17\_MACEU (1-153)  
 QY 136 ATGTGATTCATTCCTCCACACCCACTACCGAATTCACAGGATTTGGGAATCTTCTT 195  
 Db 1 MetSerileProPheSerAsnThrHisTyArgIleProGlnGlyPheGlyAsnLeuLeu 20  
 QY 196 GAAGGCTCAGCGGAGATTCGTGAGAGACCAACCGGACAAATATACCACTTTTGCAGCA 255  
 Db 21 GluGlyLeuThrArgGluileuArgGluGlnProAspAsnIleProAlaPheAla 40  
 QY 256 GCCTATTTTGAG 267  
 Db 41 AlaTyPheGlu 44

RESULT 9  
 SP17\_MONDO  
 ID SP17\_MONDO STANDARD; PRT; 179 AA.  
 AC O62771;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 05-JUL-2004 (Rel. 44, Last sequence update)  
 DE Sperm surface protein Sp17 (Sperm autoantigenic protein 17).  
 GN Names-SPA17; Synonyms-Sp17;  
 OS Monodelphis domestica (Short-tailed grey opossum).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Monodelphis.  
 OX NCBI\_TaxID=13616;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Wen Y., O'Rand M.G.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Sperm surface zona pellucida binding protein. Helps to  
 CC bind spermatozoa to the zona pellucida with high affinity. Might  
 CC function in binding zona pellucida and carbohydrates (By  
 CC similarity).  
 CC -!- SUBUNIT: Homodimer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Membrane-associated (Potential).  
 CC -!- TISSUE SPECIFICITY: Testis- and sperm-specific (By similarity).  
 CC -!- SIMILARITY: Contains 1 IQ domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AF054290; AAC08025.1; -.  
 DR HSSP; P12367; 1R2A.  
 DR InterPro; IPR000048; IQ region.  
 DR InterPro; IPR003117; RIIa.

DR Pfam; PF00612; IQ; 1.  
 DR Pfam; PF02197; RIIa; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00394; RIIa; 1.  
 DR PROSITE; PSS0096; IQ; 1.  
 KW Membrane.  
 FT DOMAIN 124 132 Poly-Glu.  
 FT DOMAIN 143 172 IQ.  
 SQ SEQUENCE 179 AA; 20864 MW; CADDCl3CEC66A00E CRC64;  
 Alignment Scores:  
 Pred. No.: 7 27e-36 Length: 179  
 Score: 44.00 Matches: 44  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 14.33% Indels: 0  
 DB: 1 Gaps: 0

AF334735 (1-954) x SP17\_MONDO (1-179)  
 QY 136 ATGTGATTCATTCCTCCACACCCACTACCGAATTCACAGGATTTGGGAATCTTCTT 195  
 Db 1 MetSerileProPheSerAsnThrHisTyArgIleProGlnGlyPheGlyAsnLeuLeu 20  
 QY 196 GAAGGCTCAGCGGAGATTCGTGAGAGACCAACCGGACAAATATACCACTTTTGCAGCA 255  
 Db 21 GluGlyLeuThrArgGluileuArgGluGlnProAspAsnIleProAlaPheAla 40  
 QY 256 GCCTATTTTGAG 267  
 Db 41 AlaTyPheGlu 44

RESULT 10  
 Q9TU08  
 ID Q9TU08 PRELIMINARY; PRT; 147 AA.  
 AC Q9TU08;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE SP17 protein.  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Caprinae; Ovis.  
 OX NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Jejunal Peyer's patch;  
 RX MEDLINE=21015023; PubMed=11132159; DOI=10.1007/s002510000253;  
 RA Tatlow D., Brownlie R., Babiuk L.A., Griebel P.;  
 RT "Differential display analysis of gene expression during the induction  
 RT of mucosal immunity.";  
 RL Immunogenetics 52:73-80 (2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Jejunal Peyer's patch;  
 RA Tatlow D., Brownlie R., Babiuk L.A., Griebel P.J.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF179926; AAD54394.1; -.  
 DR HSSP; P12367; 1R2A.  
 DR GO; GO:0008603; F:camp-dependent protein kinase regulator act. . ; IEA.  
 DR GO; GO:0007165; P:signal transduction; IEA.  
 DR InterPro; IPR000048; IQ region.  
 DR InterPro; IPR003117; RIIa.  
 DR Pfam; PF00612; IQ; 1.  
 DR Pfam; PF02197; RIIa; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00394; RIIa; 1.  
 DR PROSITE; PSS0096; IQ; 1.  
 SQ SEQUENCE 147 AA; 16903 MW; 4F9CB44103C25E66 CRC64;  
 Alignment Scores:  
 Pred. No.: 1 05e-31 Length: 147

```

Score: 40.00 Mates: 40
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 13.03% Indels: 0
DB: 2 Gaps: 0

AF334735 (1-954) x Q9TU08 (1-147)
QY 148 TTCTCAACACCCACTACCGAATTCACAGGATTTGGGAATCTTCTTGAGGGCTGACA 207
Db 5 PheSerAenThrHisTyArgIleProGlnGlyPheGlyAenLeuLeuGluGlyLeuThr 24
QY 208 CGCGAGATTTGAGAGACGACCGGACATATACCGCTTTTCAGCAGGCTATTGAG 267
Db 25 ArgGluLeuLeuArgGluGlnProAspAsnIleProAlaPheAlaAlaIlePheGlu 44

RESULT 11
Q9XT29 PRELIMINARY; PRT; 212 AA.
AC Q9XT29;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Sperm protein Sp17.
OS Callithrix jacchus (Common marmoset).
OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callitrich.
OX NCBI_TaxID=9483;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Lea I.A., O'Rand M.G.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF134585; AAD30113.1; -.
DR HSSP; P12367; 1R2A.
DR GO; GO:0008503; P: cAMP-dependent protein kinase regulator act. . . ; IEA.
DR GO; GO:0007165; P: signal transduction; IEA.
DR InterPro; IPR000048; IQ region.
DR Pfam; PF00612; IQ; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00394; RIIA; 1.
DR PROSITE; PS50096; IQ; 1.
SQ SEQUENCE 212 AA; 24304 MW; 7DCB36B0365D4CE4 CRC64;

Alignment Scores:
Pred. No.: 1.82e-24 Length: 212
Score: 33.00 Mates: 212
Percent Similarity: 100.00% Conservative: 33
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.75% Indels: 0
DB: 2 Gaps: 0

AF334735 (1-954) x Q9XT29 (1-212)
QY 166 CGAATTCACAGGATTTGGGAATCTTCTTGAGGGCTGACAGCGAGATTTCTGAGAGAG 225
Db 11 ArgIleProGlnGlyPheGlyAenLeuLeuGluGlyLeuThrArgGluLeuArgGlu 30
QY 226 CAACCGGACAAATATACCGCTTTTGAGGCTTTTCAGCAGGCTATTGAG 264
Db 31 GlnProAspAsnIleProAlaPheAlaAlaIlePhe 43

RESULT 12
Q6SJ92 PRELIMINARY; PRT; 141 AA.
AC Q6SJ92;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Sperm autoantigenic protein 17.
GN Name=SPA17;

OS Ornithorhynchus anatinus (Duckbill platypus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Monotremata; Ornithorhynchidae; Ornithorhynchus.
OX NCBI_TaxID=9258;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=21308450; PubMed=11415432; DOI=10.1042/0264-6021:3570025;
RA Wen Y., Richardson R.T., Widgren E.E., O'Rand M.G.;
RT "Characterization of Sp17: a ubiquitous three domain protein that binds heparin.";
RL Biochem. J. 357:25-31(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Alekseev O. Jr., Richardson R.T., O'Rand M.G.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY457927; AAR21214.1; -.
DR GO; GO:0008603; P: cAMP-dependent protein kinase regulator act. . . ; IEA.
DR GO; GO:0007165; P: signal transduction; IEA.
DR InterPro; IPR000048; IQ region.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF02197; RIIA; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00394; RIIA; 1.
DR PROSITE; PS50096; IQ; 1.
SQ SEQUENCE 141 AA; 15752 MW; 98E3282E0314DOB0 CRC64;

Alignment Scores:
Pred. No.: 8.95e-06 Length: 141
Score: 15.00 Mates: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.89% Indels: 0
DB: 2 Gaps: 0

AF334735 (1-954) x Q6SJ92 (1-141)
QY 223 GAGCAACGACAAATATACCGCTTTTGAGGCTTTTCAGCAGGCTATTGAG 267
Db 30 GluGlnProAspAsnIleProAlaPheAlaAlaIlePheGlu 44

RESULT 13
Q6DIJ9 PRELIMINARY; PRT; 495 AA.
AC Q6DIJ9;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Spal7-prov protein.
GN Name=spal7-prov;
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8364;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton B., Kettman M., Madan A., Rodriguez S., Sanchez A.,

```



RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smalilov D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RV [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RC Klein S., Gerhard D.S.;  
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC075539; AH75539.1; -;  
 DR GO; GO:0008603; F:AMP-dependent protein kinase regulator act. . ; IEA.  
 DR GO; GO:0007165; P:signal transduction; IEA.  
 DR InterPro; IPR000048; IQ region.  
 DR InterPro; IPR003117; RIIA.  
 DR Pfam; PF00612; IQ; 2.  
 DR Pfam; PF02197; RIIA; 1.  
 DR SMART; SM00015; IQ; 2.  
 DR SMART; SM00394; RIIA; 1.  
 DR PROSITE; PS0096; IQ; 2.  
 SQ SEQUENCE 495 AA; 56219 MW; 4B857410B790912A CRC64;

Alignment Scores:  
 Pred. No.: 0.000893 Length: 495  
 Score: 13.00 Matches: 13  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 4.23% Indels: 0  
 DB: 2 Gaps: 0

AF334735 (1-954) x Q6DIJ9 (1-495)

QY 136 ATGTGATTCCTTCACACCCACTACCGAATTC 174  
 |||||  
 Db 1 MetSerIleProPheSerAsnThrHisTyArgIlePro 13

RESULT 14

Q7S3H9 PRELIMINARY; PRT; 1012 AA.  
 AC Q7S3H9;  
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Hypothetical protein.  
 GN Name=NCU06874.1;  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OR74A;  
 RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,  
 RA Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,  
 RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,  
 RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,  
 RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,  
 RA Kothke G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,  
 RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,  
 RA Kamal M., Kamysseilis M., Mauceli E., Bielke C., Rudd S., Frishman D.,  
 RA Krystofova S., Raemissen C., Metzenberg R.L., Perkins D.D., Kroken S.,  
 RA Cognigni C., Macino G., Catcheside D., Li W., Pratt R.J., Osmani S.A.,  
 RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,  
 RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,  
 RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbold D.J., Freitag M.,  
 RA Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.;  
 RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";  
 RL Nature 0:0-0(2003).  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is

CC preliminary data.  
 DR EMBL; AABX01000409; EAA29985.1; -;  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR000910; HMG 12\_box.  
 DR Pfam; PF00505; HMG box; 1.  
 DR PROSITE; PS50118; HMG\_BOX\_2; 1.  
 DR Hypothetical protein.  
 SQ SEQUENCE 1012 AA; 110158 MW; 56323711A39D7B01 CRC64;

Alignment Scores:

Pred. No.: 0.0959 Length: 1012  
 Score: 11.00 Matches: 11  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.58% Indels: 0  
 DB: 2 Gaps: 0

AF334735 (1-954) x Q7S3H9 (1-1012)

QY 920 TCTATCATCCCAAAAAAAAAAAAAAAAAAAAA 952  
 |||||  
 Db 581 SerIleThrSerProLysLysLysLysLys 591

RESULT 15

Q8FN75 PRELIMINARY; PRT; 1148 AA.  
 AC Q8FN75;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Putative ribonuclease.  
 GN OrderedLocustNames=CE2272;  
 OS Corynebacterium efficiens.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 OX NCBI\_TaxID=152794;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=YS-314;  
 RX MEDLINE=2273752; PubMed=12840036; DOI=10.1101/gr.1285603;  
 RA Nishio Y., Nakamura Y., Kawarabayashi Y., Usuda Y., Kimura E.,  
 RA Sugimoto S., Matsui K., Yamagishi A., Kikuchi H., Ikeo K.,  
 RA Gojobori T.;  
 RT "Comparative complete genome sequence analysis of the amino acid  
 RT replacements responsible for the thermostability of Corynebacterium  
 RT efficiens.";  
 RL Genome Res. 13:1572-1579(2003).  
 DR EMBL; AP005221; BAC19082.1; -;  
 DR GO; GO:0005737; C:cytoplasm; IEA.  
 DR GO; GO:0004540; F:ribonuclease activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003743; F:translation initiation factor activity; IEA.  
 DR GO; GO:0006396; P:RNA processing; IEA.  
 DR GO; GO:0006413; P:translational initiation; IEA.  
 DR InterPro; IPR006847; IP2\_N.  
 DR InterPro; IPR008994; Nucleic acid OB.  
 DR InterPro; IPR009061; Putativ\_DNA\_Bind.  
 DR InterPro; IPR004659; RNaseEG.  
 DR InterPro; IPR003029; S1.  
 DR Pfam; PF04760; IP2\_N; 1.  
 DR Pfam; PF00575; S1; 1.  
 DR SMART; SM00316; S1; 1.  
 DR TIGRFAMs; TIGR00757; RNaseEG; 1.  
 DR PROSITE; PS50126; S1; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 1148 AA; 126647 MW; C364493697D4DC37 CRC64;

Alignment Scores:

Pred. No.: 1.03 Length: 1148  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0

Query Match: 3.26% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x Q8FN75 (1-1148)

QY 65 GAAACACACCGGACCGCGCGCAGCT 94  
|||  
Db 1124 GlulysGlnProGluProAlaProAla 1133

Search completed: February 15, 2005, 02:28:53  
Job time : 207 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 15, 2005, 02:13:41 ; Search time 45 Seconds  
(without alignments)  
4079.591 Million cell updates/sec

Title: AF334735  
Perfect score: 307  
Sequence: 1 TGGCCCTCTCGCGCGCG.....AAAAAAAAAAAAAAAAAAAA 954

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues  
Word size: 1

Total number of hits satisfying chosen parameters: 565918  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DBV=xlp  
-O=/cgn2.1/USPTO.spool\_p/af334735/runat.14022005.160238.19349/app\_query.fasta\_1.1095  
-DB=PIR 79 -QFMT=fastan -SUFFIX=oligo.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -MIST=45  
-DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=AF334735 @CGN 1.1.63 @runat.14022005.160238.19349 -NCPV=6 -ICPV=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DBV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DSLOP=6 -DELEXT=7

Database : PIR 79:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	151	49.2	151	2 I38243	zona pellucida bin
2	44	14.3	146	2 I46506	zona pellucida bin
3	44	14.3	149	2 S49527	zona pellucida bin
4	9	2.9	145	2 I51706	stathmin - African
5	9	2.9	296	2 T20726	hypothetical prote
6	8	2.6	41	2 T07274	hypothetical prote
7	8	2.6	48	2 A26334	beta-casein - wate
8	8	2.6	139	2 S30850	hypothetical prote
9	8	2.6	151	2 C90082	hypothetical prote
10	8	2.6	151	2 B90102	hypothetical prote
11	8	2.6	151	2 P90101	hypothetical prote
12	8	2.6	151	2 C90118	hypothetical prote
13	8	2.6	151	2 A90138	hypothetical prote
14	8	2.6	209	2 A59068	beta-casein varian

C 15	8	2.6	222	2 JCI384	beta-casein precu
C 16	8	2.6	222	2 A32979	beta-casein precu
C 17	8	2.6	224	1 KBBOA2	beta-casein precu
C 18	8	2.6	226	1 KBHU	embryogenesis-rela
C 19	8	2.6	226	2 T09567	beta-casein - pig
C 20	8	2.6	232	2 A48384	homotetic protein T
C 21	8	2.6	232	2 A61045	hypothetical prote
C 22	8	2.6	250	2 S19518	cold acclimation p
C 23	8	2.6	262	2 S43953	hypothetical prote
C 24	8	2.6	296	2 E71118	NADH dehydrogenase
C 25	8	2.6	349	2 T12120	hypothetical prote
C 26	8	2.6	368	2 G85587	probable inner mem
C 27	8	2.6	368	2 AD0599	hypothetical prote
C 28	8	2.6	368	2 F90737	hypothetical prote
C 29	8	2.6	368	2 H64815	ybhr protein - Esc
C 30	8	2.6	393	2 T33715	hypothetical prote
C 31	8	2.6	422	2 C70474	Na+/H+-exchanging
C 32	8	2.6	448	2 T30982	hypothetical prote
C 33	8	2.6	493	2 F96966	protein Flm21.12 [
C 34	8	2.6	517	2 T49310	hypothetical prote
C 35	8	2.6	624	2 T20445	hypothetical prote
C 36	8	2.6	629	2 T45796	hypothetical prote
C 37	8	2.6	661	2 T37753	hypothetical prote
C 38	8	2.6	869	2 A71400	probable disease r
C 39	8	2.6	897	2 T21688	hypothetical prote
C 40	8	2.6	975	2 T48107	hypothetical prote
C 41	8	2.6	1091	1 PWB9R2	Ca2+-transporting
C 42	8	2.6	1091	2 S67852	Ca2+-transporting
C 43	8	2.6	1091	2 S25007	Ca2+-transporting
C 44	8	2.6	1192	2 A71623	probable secreted
C 45	8	2.6	1234	2 C97606	hypothetical prote

ALIGNMENTS

RESULT 1

I38243  
zona pellucida binding protein Sp17 [similarity] - human  
C:Species: Homo sapiens (man)  
C>Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 09-Jul-2004  
C:Accession: I38243; S52921  
R:Richardson, R.T.; Yamasaki, N.; O'Rand, M.G.  
Dev. Biol. 165, 688-701, 1994  
A:Title: Sequence of a rabbit sperm zona pellucida binding protein and localization dur  
A:Reference number: I38243; MUID:95046885; PMID:7525387  
A:Accession: I38243  
A>Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-151 <RES>  
A:Cross-references: UNIPROT:Q15506; EMBL:Z48570; NID:G695580; PIDN:CAA8459.1; PID:G695  
R:Lea, I.A.; Richardson, R.T.; Widgren, E.E.; O'Rand, M.G.  
Submitted to the EMBL Data Library, March 1995  
A:Description: Human Sp17: a sperm-zona binding protein.  
A:Reference number: S52921  
A:Accession: S52921  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <LEA>  
C:Superfamily: sperm surface protein Sp17

Alignment Scores:  
Pred. No.: 4.44e-145 Length: 151  
Score: 151.00 Matches: 151  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 49.19% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x I38243 (1-151)

QY 136 ATGTGATTCATTCCTCCACACCCACTACCGAATTCACAGGATTGGGATCTCTT 195  
|||||  
Db 1 MetSerIleProPheSerAsnThrHisTyrArgIleProGlnGlyPheGlyAsnLeuLeu 20  
|||||

QY 196 GAAGGCTGACACGGAGATTCTGAGAGAGCAACCGGCAATATATACAGCTTTTGCAGCA 255  
Db 21 GluGlyLeuThrArgGluLeuLeuArgGluGlnProAspAsnIleProAlaPheAlaA 40  
QY 256 GCTATTTTGGAGCGCTCTAGAGAAAGAGAGAAACCAACTTTTGATCCAGAGAAATGG 315  
Db 41 AlaTyPheGluSerLeuLeuGluLysArgGluLeuThrAsnPheAspProAlaGluTrp 60  
QY 316 GGGAGTAGGTAGAGACCGCTTCTATTAACAATCATGCAATTCGAGAGAGCAAGAACCT 375  
Db 61 GlySerLysValGluAspArgPheTyAsnAsnHisAlaPheGluGluGlnGluProPro 80  
QY 376 GAGAAAGTGATCTTAACAAGAGAGTCTAGATATCTGGAGAGGAGGAGAGAGCATCA 435  
Db 81 GluLysSerAspProLysGlnGluSerGlnIleSerGlyGluGluGlnGluThrSer 100  
QY 436 GTCCACATCTTACATCTTCTCAGGAAGATAAGAGAAAGAGAGGTTCTGCTGTCAA 495  
Db 101 ValThrIleLeuAspSerSerGluGluAspLysGluLysGluGluValAlaAlaValLys 120  
QY 496 ATCCAAGTCTTCCGGGACACATAGCCAGAGAGGAGGCAAGAAATGAAACAAAT 555  
Db 121 IleGlnAlaAlaPheArgGlyHisIleAlaArgGluGluAlaLysLysMetLysThrAen 140  
QY 556 ACTCTTCANAATCAGGAAAAGAGGAAACAG 588  
Db 141 SerLeuGlnAsnGluLysGluGluAsnLys 151  
RESULT 2  
I46506  
zona pellucida binding protein Sp17 [similarity] - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 14-Feb-1997 #sequence\_revision 14-Feb-1997 #text\_change 09-Jul-2004  
C:Accession: I46506; S58441; S31764  
R:Richardson, R.T.; Yamasaki, N.; O'Rand, M.G.  
Dev. Biol. 165, 688-701, 1994  
A:Title: Sequence of a rabbit sperm zona pellucida binding protein and localization dur  
A:Reference number: I38243; MUID:95046885; PMID:7525387  
A:Accession: I46506  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-146 <RIC>  
A:Cross-references: UNIPROT:P36425; EMBL:Z20655; NID:G479083; PIDN:CAA79674.1; PID:G4790  
A:Accession: S58441  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 113-126 <RI3>  
R:Richardson, R.T.; Yamasaki, N.; O'Rand, M.G.  
submitted to the EMBL Data Library, January 1993  
A:Description: Sequence of a mammalian sperm surface zona pellucida binding protein and  
A:Reference number: S31764  
A:Accession: S31764  
A:Molecule type: DNA  
A:Residues: 1-122, DT', <RI2>  
A:Residues: 1-122, DT', <RI2>  
A:Cross-references: EMBL:Z20655  
C:Superfamily: sperm surface protein Sp17

Alignment Scores:  
Pred. No.: 6.32e-36 Length: 146  
Score: 44.00 Matches: 44  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 14.33% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x I46506 (1-146)

QY 136 ATGTCGATTCCATTTCCACACCCCACTACCGAATTCACAGGATTTGGGAATCTTCT 195  
Db 1 MetSerIleProPheSerAsnThrHisTyArgIleProGlnGlyPheGlyAsnLeuLeu 20  
QY 196 GAAGGCTGACACGGAGATTCTGAGAGAGCAACCGGCAATATATACAGCTTTTGCAGCA 255

Db 21 GluGlyLeuThrArgGluLeuLeuArgGluGlnProAspAsnIleProAlaPheAlaA 40  
QY 256 GCTATTTTGGAG 267  
Db 41 AlaTyPheGlu 44  
RESULT 3  
S49527  
zona pellucida binding protein Sp17 [similarity] - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004  
C:Accession: S49527  
R:Kong, M.; Richardson, R.T.; O'Rand, M.G.  
submitted to the EMBL Data Library, October 1994  
A:Description: Sequence and localization of the mouse sperm protein, Sp17.  
A:Reference number: S49527  
A:Accession: S49527  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-149 <RON>  
A:Cross-references: UNIPROT:Q62252; EMBL:Z46299; NID:G561527; PIDN:CAA86455.1; PID:G5615  
C:Superfamily: sperm surface protein Sp17

Alignment Scores:  
Pred. No.: 6.31e-36 Length: 149  
Score: 44.00 Matches: 44  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 14.33% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x S49527 (1-149)

QY 136 ATGTCGATTCCATTTCCACACCCCACTACCGAATTCACAGGATTTGGGAATCTTCT 195  
Db 1 MetSerIleProPheSerAsnThrHisTyArgIleProGlnGlyPheGlyAsnLeuLeu 20  
QY 196 GAAGGCTGACACGGAGATTCTGAGAGAGCAACCGGCAATATATACAGCTTTTGCAGCA 255  
Db 21 GluGlyLeuThrArgGluLeuLeuArgGluGlnProAspAsnIleProAlaPheAlaA 40  
QY 256 GCTATTTTGGAG 267  
Db 41 AlaTyPheGlu 44

RESULT 4  
I51706  
stathmin - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jul-2004  
C:Accession: I51706  
R:Maucuer, A.; Moreau, J.; Mechali, M.; Sobel, A.  
J. Biol. Chem. 268, 16420-16429, 1993  
A:Title: Stathmin gene family: phylogenetic conservation and developmental regulation in  
A:Reference number: A47345; MUID:93346387; PMID:8344928  
A:Accession: I51706  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-145 <MAU>  
A:Cross-references: UNIPROT:Q09006; EMBL:X71431; NID:G397173; PIDN:CAA50562.1; PID:G3971  
C:Genetics:  
A:Gene: XO35  
C:Superfamily: stathmin

Alignment Scores:  
Pred. No.: 3.2 Length: 145  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.93% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x I51706 (1-145)

QY 917 CTCTATCATCCCTCCCAAAAAA 943  
|||||  
Db 35 LeuSerIleThrSerProLysLysLys 43

## RESULT 5

T20726  
hypothetical protein F10G8.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T20726  
R:Basham, V.  
submitted to the EMBL Data Library, September 1996  
A:Reference number: Z19315  
A:Accession: T20726  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-296 <WIL>  
A:Cross-references: UNIPROT:Q93456; EMBL:Z80216; PIDN:GN00019; CESP:F1  
A:Experimental source: clone F10G8  
C:Genetics:  
A:Gene: CESP:F10G8.7  
A:Map position: 1  
A:Introns: 53/1; 129/3; 153/3; 187/3; 209/2; 249/2

## Alignment Scores:

Pred. No.:	2.94	Length:	296
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.92%	Indels:	0
DB:	2	Gaps:	0

AF334735 (1-954) x T20726 (1-296)

QY 482 ACTCTCTTTTCCTATCTCTCA 456  
|||||  
Db 27 ThrSerSerPheSerLeuSerSer 35

## RESULT 6

T07274  
hypothetical protein 41c - Chlorella vulgaris chloroplast  
C:Species: chloroplast Chlorella vulgaris  
C>Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 21-Jul-2000  
C:Accession: T07274  
R:Wakasugi, T.; Nagai, T.; Kapoor, M.; Sugita, M.; Ito, M.; Ito, S.; Tsudzuki, J.; Nakas  
Proc. Natl. Acad. Sci. U.S.A. 94, 5967-5972, 1997  
A:Title: Complete nucleotide sequence of the chloroplast genome from the green alga Chlo  
A:Reference number: Z15985; MUID:97303241; PMID:9159184  
A:Accession: T07274  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-41 <WAK>  
A:Cross-references: EMBL:AB001684; NID:g2224352; PIDN:BAA57922.1; PID:g2224438  
C:Genetics:  
A:Genome: chloroplast  
C:Keywords: chloroplast

Alignment Scores:			
Pred. No.:	38.9	Length:	41
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.61%	Indels:	0
DB:	2	Gaps:	0

AF334735 (1-954) x T07274 (1-41)

QY 24 TTTTCTTTTCTTAAGAAAAA 47  
|||||  
Db 6 PhePhePhePheLeuLysLysLys 13

## RESULT 7

A26334  
beta-casein - water buffalo (fragment)  
C:Species: Bubalus arnee (water buffalo)  
C>Date: 19-Nov-1988 #sequence\_revision 19-Nov-1988 #text\_change 09-Jul-2004  
C:Accession: A26334  
R:Petrilli, P.; Pucci, P.; Morris, H.R.; Addeo, F.  
Biochem. Biophys. Res. Commun. 140, 28-37, 1986  
A:Title: Assignment of phosphorylation sites in buffalo beta-casein by fast atom bombar  
A:Reference number: A26334; MUID:87048757; PMID:3778448  
A:Accession: A26334  
A:Molecule type: protein  
A:Residues: 1-48 <PET>  
A:Cross-references: UNIPROT:Q7M2U5  
C:Superfamily: beta-casein

Alignment Scores:			
Pred. No.:	38.2	Length:	48
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.60%	Indels:	0
DB:	2	Gaps:	0

AF334735 (1-954) x A26334 (1-48)

QY 470 TCCTTATCTCTCAGAGAGTCT 447  
|||||

Db 15 SerLeuSerSerSerGluGluSer 22

## RESULT 8

S30850  
hypothetical protein YEL033w - yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C>Date: 28-May-1993 #sequence\_revision 28-May-1993 #text\_change 09-Jul-2004  
C:Accession: S30850; S50511  
R:Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Schl, P.; Komp, C.; Wei, Y.; Taylor,  
submitted to the EMBL Data Library, February 1993  
A:Reference number: S30812  
A:Accession: S30850  
A:Molecule type: DNA  
A:Residues: 1-139 <MUL>  
A:Cross-references: UNIPROT:P32633; GB:U18779; EMBL:L10830; NID:g603625; PIDN:AAB65009.1

R:Dietrich, F.S.  
submitted to the EMBL Data Library, December 1994  
A:Description: The sequence of S. cerevisiae cosmids 8199, 8334, and 9871.  
A:Reference number: S50491  
A:Accession: S50511  
A:Molecule type: DNA  
A:Residues: 1-139 <DIE>  
A:Cross-references: EMBL:U18779; NID:g603625; PIDN:AAB65009.1; PID:g603646; MIPS:YEL033w

C:Genetics:  
A:Cross-references: SGD:S0000759  
A:Map position: 5L  
C:Superfamily: Saccharomyces hypothetical protein YEL033w  
C:Keywords: transmembrane protein

Alignment Scores:			
Pred. No.:	33.7	Length:	139
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.61%	Indels:	0
DB:	2	Gaps:	0

AF334735 (1-954) x S30850 (1-139)

QY 931 CCCCCCAAAAAAAAAAAAAA 954  
|||||

Db 103 ProGlnLysLysLysLysLys 110

## RESULT 9

C90082

hypothetical protein orf151 [imported] - Guillardia theta nucleomorph  
C:Species: nucleomorph Guillardia theta  
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: C90082  
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rei  
Nature 410, 1091-1096, 2001  
A:Title: The highly reduced genome of an enslaved algal nucleus.  
A:Reference number: A99082; PMID:11323671; PMID:11323671  
A:Accession: C90082  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <DOU>  
A:Cross-references: UNIPROT:Q9AQY4; GB:AF165818; NID:G13794566; PIDN:AAK39941.1; GSPDB:G  
C:Genetics:  
A:Gene: orf151  
A:Map position: 1  
A:Genome: nucleomorph  
C:Keywords: nucleomorph

Alignment Scores:  
Pred. No.: 33.4 Length: 151  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.60% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x C90082 (1-151)

QY 594 TCCTCACTGTTTCTCTCTTTTC 571

Db 28 SerSerLeuValPheLeuPhePhe 35

RESULT 10

B90102  
hypothetical protein orf151 [imported] - Guillardia theta nucleomorph  
C:Species: nucleomorph Guillardia theta  
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: B90102  
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rei  
Nature 410, 1091-1096, 2001  
A:Title: The highly reduced genome of an enslaved algal nucleus.  
A:Reference number: A99082; PMID:11323671; PMID:11323671  
A:Accession: B90102  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <DOU>  
A:Cross-references: UNIPROT:Q9AQY4; GB:AJ010592; NID:G12580657; PIDN:CAC26974.1; GSPDB:G

C:Genetics:

A:Gene: orf151

A:Map position: 2

A:Genome: nucleomorph

C:Keywords: nucleomorph

Alignment Scores:  
Pred. No.: 33.4 Length: 151  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.60% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x B90102 (1-151)

QY 594 TCCTCACTGTTTCTCTCTTTTC 571

Db 28 SerSerLeuValPheLeuPhePhe 35

RESULT 11

F90101  
hypothetical protein orf151b [imported] - Guillardia theta nucleomorph

C:Species: nucleomorph Guillardia theta  
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: F90101  
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rei  
Nature 410, 1091-1096, 2001  
A:Title: The highly reduced genome of an enslaved algal nucleus.  
A:Reference number: A99082; PMID:11323671; PMID:11323671  
A:Accession: F90101  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <DOU>  
A:Cross-references: UNIPROT:Q9AQY4; GB:AF165818; NID:G13794573; PIDN:AAK39948.1; GSPDB:G  
C:Genetics:  
A:Gene: orf151b  
A:Map position: 1  
A:Genome: nucleomorph  
C:Keywords: nucleomorph

Alignment Scores:  
Pred. No.: 33.4 Length: 151  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.60% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x F90101 (1-151)

QY 594 TCCTCACTGTTTCTCTCTTTTC 571

Db 28 SerSerLeuValPheLeuPhePhe 35

RESULT 12

C90118  
hypothetical protein orf151 [imported] - Guillardia theta nucleomorph  
C:Species: nucleomorph Guillardia theta  
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: C90118  
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rei  
Nature 410, 1091-1096, 2001  
A:Title: The highly reduced genome of an enslaved algal nucleus.  
A:Reference number: A99082; PMID:11323671; PMID:11323671  
A:Accession: C90118  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <DOU>  
A:Cross-references: UNIPROT:Q9AQY4; GB:AF083031; NID:G13794413; PIDN:AAK39790.1; GSPDB:G

C:Genetics:

A:Gene: orf151

A:Map position: 3

A:Genome: nucleomorph

C:Keywords: nucleomorph

Alignment Scores:  
Pred. No.: 33.4 Length: 151  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.60% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x C90118 (1-151)

QY 594 TCCTCACTGTTTCTCTCTTTTC 571

Db 28 SerSerLeuValPheLeuPhePhe 35

RESULT 13

A90138  
hypothetical protein orf151b [imported] - Guillardia theta nucleomorph  
C:Species: nucleomorph Guillardia theta



A;Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C;Accession: A90138  
R;Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Reid,  
Nature 410, 1091-1096, 2001  
A;Title: The highly reduced genome of an enslaved algal nucleus.  
A;Reference number: A99082; MUID:11323671; PMID:11323671  
A;Accession: A90138  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-151 <DOU>  
A;Cross-references: UNIPROT:Q9A0Y4; GB:AF083031; NID:gl3794420; PIDN:AAK39797.1; GSPDB:G  
C;Genetics:  
A;Gene: orf15b  
A;Map position: 3  
A;Genome: nucleomorph  
C;Keywords: nucleomorph

Alignment Scores:  
Pred. No.: 33.4 Length: 151  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.60% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x A90138 (1-151)

QY 594 TCCTCACTGTTTCCTCTTTTC 571

Db 28 SerSerLeuValPheLeuPhePhe 35

#### RESULT 14

A59068

beta-casein variant CnH - bovine  
C;Species: Bos primigenius taurus (cattle)  
C;Date: 20-Sep-1999 #sequence\_revision 24-Sep-1999 #text\_change 24-Sep-1999  
C;Accession: A59068; B59068  
R;Han, S.K.; Shin, Y.C.  
Anim. Genet. 27(Suppl.2), 91b, 1996  
A;Title: Biochemical characterization of the new beta-casein variant in Korean cattle.  
A;Reference number: A59068  
A;Accession: A59068  
A;Status: protein sequence not shown  
A;Molecule type: protein  
A;Residues: 1-209 <HAN>  
A;Experimental source: strain Korean cattle  
A;Note: submitted to the Protein Sequence Database, September 1999  
A;Note: includes casein phosphopeptide H  
A;Accession: B59068  
A;Status: protein sequence not shown  
A;Molecule type: protein  
A;Residues: 1-28 <HAN2>  
A;Experimental source: strain Korean cattle  
C;Superfamily: beta-casein  
C;Keywords: milk; phosphoprotein  
F;15,17,18,19/Binding site: phosphate (Ser) (covalent) #status predicted

Alignment Scores:  
Pred. No.: 32.1 Length: 209  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.60% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x A59068 (1-209)

QY 470 TCCTATCTTCCTCAGAGAGTCT 447

Db 15 SerLeuSerSerSerGluGluSer 22

#### RESULT 15

JC1384

beta-casein precursor - goat  
C;Species: Capra aegagrus hircus (domestic goat)  
C;Date: 10-Jun-1993 #sequence\_revision 10-Jun-1993 #text\_change 09-Jul-2004  
C;Accession: JC1384  
R;Roberts, B.; DiTullio, P.; Vitale, J.; Hehir, K.; Gordon, K.  
Gene 121, 255-262, 1992  
A;Title: Cloning of the goat beta-casein-encoding gene and expression in transgenic mic  
A;Reference number: JC1384; MUID:93077039; PMID:1446822  
A;Accession: JC1384  
A;Molecule type: DNA  
A;Residues: 1-222 <ROB>  
A;Cross-references: UNIPROT:P33048; GB:M90556  
C;Genetics:  
A;Gene: CSN2  
A;Introns: 17/3; 26/3; 35/3; 45/3; 57/3; 221/3  
C;Superfamily: beta-casein  
C;Keywords: milk; phosphoprotein

Alignment Scores:  
Pred. No.: 31.9 Length: 222  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.60% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x JC1384 (1-222)

QY 470 TCCTATCTTCCTCAGAGAGTCT 447

Db 30 SerLeuSerSerSerGluGluSer 37

Search completed: February 15, 2005, 02:30:27  
Job time : 48 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 15, 2005, 02:04:41 ; Search time 168 Seconds  
(without alignments)  
4392.493 Million cell updates/sec

Title: AF334735  
Perfect score: 307  
Sequence: 1 TCGCCCTTCGCGCGCG.....AAAAAAAAAAAAAAAAAAAA 954

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Word size: 1

Total number of hits satisfying chosen parameters: 3967864

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool\_p/AF334735/runat\_14022005\_160237\_19333/app.query.fasta\_1.1095  
-DB=A\_Geneseq\_16Dec04 -Qfmt=fastan -SUFFIX=oligo.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=AF334735 @CGN 1 1 224 @runat\_14022005\_160237\_19333 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DESBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : A\_Geneseq\_16Dec04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	151	49.2	151	2 AAR79761	Aar79761 Human tes
2	151	49.2	151	2 AAW42642	Aaw42642 Human spe
3	151	49.2	162	2 AAW42692	Aaw42692 Recombina
4	76	24.8	119	3 AAG00833	Aag00833 Human sec
5	76	24.8	140	4 ABG11758	Abg11758 Novel hum
6	57	18.6	300	4 ABG08152	Abg08152 Novel hum
7	43	14.0	163	2 AAW42693	Aaw42693 Baboon sp
8	33	10.7	146	2 AAW42691	Aaw42691 Recombina
9	28	9.1	28	2 AAW42767	Aaw42767 Peptide o
10	25	8.1	25	2 AAW42762	Aaw42762 Peptide o

11	25	8.1	273	4 ABG11759	Abg11759 Novel hum
12	21	6.8	21	2 AAW42771	Aaw42771 Peptide d
13	21	6.8	21	2 AAW42770	Aaw42770 Peptide o
14	21	6.8	21	2 AAW42772	Aaw42772 Peptide d
15	19	6.2	19	2 AAW42869	Aaw42869 Peptide d
16	17	5.5	17	2 AAW42773	Aaw42773 Peptide d
17	17	5.5	18	2 AAW42874	Aaw42874 Antigenic
18	16	5.2	16	2 AAW42764	Aaw42764 Peptide o
19	16	5.2	16	2 AAW42868	Aaw42868 Peptide o
20	16	5.2	16	2 AAW42873	Aaw42873 Antigenic
21	13	4.2	13	2 AAW42870	Aaw42870 Sp17 pept
22	13	4.2	21	2 AAW42768	Aaw42768 Peptide o
23	13	4.2	35	2 AAW42871	Aaw42871 Sp17 pept
24	12	3.9	28	2 AAW42766	Aaw42766 Peptide o
25	11	3.6	16	2 AAW42763	Aaw42763 Peptide o
26	10	3.3	10	2 AAR79763	Aar79763 Rabbit te
27	10	3.3	10	2 AAR79764	Aar79764 Rabbit te
28	10	3.3	10	2 AAR79767	Aar79767 Rabbit te
29	10	3.3	10	2 AAR79777	Aar79777 Rabbit te
30	10	3.3	10	2 AAR79778	Aar79778 Rabbit te
31	10	3.3	10	2 AAR79766	Aar79766 Rabbit te
32	10	3.3	10	2 AAR79779	Aar79779 Rabbit te
33	10	3.3	10	2 AAR79773	Aar79773 Rabbit te
34	10	3.3	10	2 AAR79762	Aar79762 Rabbit te
35	10	3.3	10	2 AAR79772	Aar79772 Rabbit te
36	10	3.3	10	2 AAR79768	Aar79768 Rabbit te
37	10	3.3	10	2 AAR79770	Aar79770 Rabbit te
38	10	3.3	10	2 AAR79765	Aar79765 Rabbit te
39	10	3.3	10	2 AAR79769	Aar79769 Rabbit te
40	10	3.3	10	2 AAR79771	Aar79771 Rabbit te
41	10	3.3	10	2 AAW42733	Aaw42733 Antigenic
42	10	3.3	10	2 AAW42740	Aaw42740 Antigenic
43	10	3.3	10	2 AAW42744	Aaw42744 Antigenic
44	10	3.3	10	2 AAW42779	Aaw42779 Antigenic
45	10	3.3	10	2 AAW42793	Aaw42793 Antigenic

ALIGNMENTS

RESULT 1

AAR79761

ID AAR79761 standard; protein; 151 AA.

XX AC AAR79761;

XX 11-DEC-1995 (first entry)

XX DT Human testis sperm zona binding protein Sp17.

XX DE Sperm antigen; autoantigen; zona pellucida; binding protein;

XX KW contraceptive; fertility.

XX OS Homo sapiens.

XX PN W09515764-A1.

XX PD 15-JUN-1995.

XX PF 17-NOV-1994; 94WO-US013328.

XX PR 10-DEC-1993; 93US-00166195.

XX PA (UYNC-) UNIV NORTH CAROLINA.

XX PI Orand MG, Widgren EE, Richardson RT, Lea IA;

XX DR WPI; 1995-224147/29.

XX DR N-PSDB; AAQ98173.

XX PT Novel sperm antigens useful as immunocontraceptive agents - and for

XX diagnosis of autoimmune infertility.

XX PS Claim 6; Page 30-31; 62pp; English.

XX The cloning and sequence data of rabbit Sp17 was previously reported  
 CC (Richardson and O'Rand, Mol. Biol. Cell 3, 15a (1992)). This protein is  
 CC known to be a member of the rabbit sperm antigen (RSA) family of rabbit  
 CC testis/sperm autoantigens and is also to be expressed in mice. A human  
 CC testis cDNA library was screened using the protein coding region of the  
 CC rabbit Sp17 gene as a probe. One clone contd. an insert 71% identical to  
 CC the rabbit Sp17 gene at the nt. level (see AAQ98173). This clone encoded  
 CC a protein (AAR79761) with a calc. mol. wt. of 17,534 Da, 76.7% identical  
 CC to the rabbit Sp17 and 71.8% identical to the mouse Sp17 protein sequence  
 XX

SQ Sequence 151 AA;

Alignment Scores:  
 Pred. No.: 5.3e-143 Length: 151  
 Score: 151.00 Matches: 151  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 49.19% Indels: 0  
 DB: 2 Gaps: 0

AF334735 (1-954) x AAR79761 (1-151)

QY 136 ATGTCGATTCATTCCTCCACACCCACTACCGAATTCACAGGATTGGGAATCTCTT 195  
 DB 1 MetSerIleProPheSerAsnThrHisTyrArgIleProGlnGlyPheGlyAsnLeuLeu 20  
 QY 196 GAAGGCTCACGCGAGATCTGAGAGACACCGGACCAATATACCAGCTTTTCAGCA 255  
 DB 21 GluGlyLeuThrArgGluIleLeuArgGluGlnProAspAsnIleProAlaPheAla 40  
 QY 256 GCCTATTTTGGAGCGCTTCTAGAGAAAGAGAGAAACCAACTTTTGATCCAGCAATGG 315  
 DB 41 AlaTyrPheGluSerLeuLeuGluLysArgGluLysThrAsnPheAspProAlaGluTrp 60  
 QY 316 GGGAGTAAGTAGAGACCGCTTCTATACCAATCATGCTTCGAGAGACCAAGACCACT 375  
 DB 61 GlySerLysValGluAspArgPheTyrAsnAsnHisAlaPheGluGluGlnProPro 80  
 QY 376 GAGAAAGTGCCTTAACAGACAGAGTCTCAGATATCTGGAGAGGAGAGAGACATCA 435  
 DB 81 GluLysSerAspProLysGlnGluSerGlnIleSerGlyLysGluGluGluThrSer 100  
 QY 436 GTCCACCATCTAGACTCTTCTGAGGAGACATAGCCAGAGAGGAGGAGGAGGAGGAG 495  
 DB 101 ValThrIleLeuAspSerGluGluAspLysGluLysGluLysValAlaAlaValLys 120  
 QY 496 ATCCAGCTGCTTCCGGGGACACATAGCCAGAGAGGAGGAGGAGGAGGAGGAGGAG 555  
 DB 121 IleGlnAlaAlaPheArgGlyHisIleAlaArgGluGluAlaLysLysMetLysThrAsn 140  
 QY 556 AGTCTTCAAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 588  
 DB 141 SerLeuGlnAsnGluGluLysGluGluAsnLys 151

RESULT 2

AAW42642

ID AAW42642 standard; protein; 151 AA.

XX

AC AAW42642;

XX 27-APR-1998 (first entry)

XX Human sperm zona binding protein Sp17.

XX Sperm zona binding protein Sp17; zona pellucida; binding; reduction;  
 KW Human sperm autoantigen; antigen; antigenic epitope; immunocontraception;  
 KW fertility; immunocontraceptive vaccine; autoimmune fertility.

XX Homo sapiens.

OS

XX

PN W09739020-A2.

XX

PD 23-OCT-1997.  
 XX 11-APR-1997; 97WO-US006489.  
 XX 15-APR-1996; 96US-00632535.  
 XX (UYN-C) UNIV NORTH CAROLINA.  
 XX Orand MG, Lea I, Widgren EE;  
 XX WPI; 1997-526394/48.  
 DR N-PSDB; AAV04818.  
 XX

Sperm protein Sp17 antigenic peptide(s) - used as immuno:contraceptive to  
 reduce fertility.

Example 1; Page 59; 160pp; English.

The present sequence represents a human sperm zona binding protein Sp17.  
 The cDNA sequence was identified by screening a human testis cDNA library  
 with the coding region of the rabbit Sp17 gene as a probe. The human Sp17  
 protein binds to zona pellucida at high affinity by binding sulphated,  
 complex carbohydrates. The protein has been shown to be a human sperm  
 autoantigen. Peptides comprising a contiguous segment of the amino acid  
 of Sp17 have been found to be antigenic in several species. Peptides  
 AAW42774-867 contain antigenic epitopes in mice, rabbits, non human  
 primates and humans. The peptides can be used in immunocontraceptive  
 methods to reduce the fertility of animals, particularly in females. The  
 peptides can be used in an immunocontraceptive vaccine. The presence of  
 the antigenic peptides can be detected to diagnose autoimmune fertility  
 in both male and female subjects

SQ Sequence 151 AA;

Alignment Scores:

Pred. No.: 5.3e-143 Length: 151  
 Score: 151.00 Matches: 151  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 49.19% Indels: 0  
 DB: 2 Gaps: 0

AF334735 (1-954) x AAW42642 (1-151)

QY 136 ATGTCGATTCATTCCTCCACACCCACTACCGAATTCACAGGATTGGGAATCTCTT 195  
 DB 1 MetSerIleProPheSerAsnThrHisTyrArgIleProGlnGlyPheGlyAsnLeuLeu 20  
 QY 196 GAAGGCTCACGCGAGATCTGAGAGACACCGGACCAATATACCAGCTTTTCAGCA 255  
 DB 21 GluGlyLeuThrArgGluIleLeuArgGluGlnProAspAsnIleProAlaPheAla 40  
 QY 256 GCCTATTTTGGAGCGCTTCTAGAGAAAGAGAGAAACCAACTTTTGATCCAGCAATGG 315  
 DB 41 AlaTyrPheGluSerLeuLeuGluLysArgGluLysThrAsnPheAspProAlaGluTrp 60  
 QY 316 GGGAGTAAGTAGAGACCGCTTCTATACCAATCATGCTTCGAGAGACCAAGACCACT 375  
 DB 61 GlySerLysValGluAspArgPheTyrAsnAsnHisAlaPheGluGluGlnProPro 80  
 QY 376 GAGAAAGTGCCTTAACAGACAGAGTCTCAGATATCTGGAGAGGAGAGAGGAGGAG 435  
 DB 81 GluLysSerAspProLysGlnGluSerGlnIleSerGlyLysGluGluGluThrSer 100  
 QY 436 GTCCACCATCTAGACTCTTCTGAGGAGACATAGCCAGAGAGGAGGAGGAGGAGGAG 495  
 DB 101 ValThrIleLeuAspSerGluGluAspLysGluLysGluLysValAlaAlaValLys 120  
 QY 496 ATCCAGCTGCTTCCGGGGACACATAGCCAGAGAGGAGGAGGAGGAGGAGGAGGAG 555  
 DB 121 IleGlnAlaAlaPheArgGlyHisIleAlaArgGluGluAlaLysLysMetLysThrAsn 140  
 QY 556 AGTCTTCAAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 588

[illegible]

The present sequence is a polypeptide encoded by one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were prepared from total human RNAs or poly(A) RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.

XX activity of (II) or to treat disease states involving (II). (II) is  
SQ useful for generating antibodies against it, detecting or quantitating a  
polypeptide in tissue, as molecular weight markers and as a food  
supplement. (II) and its binding partners are useful in medical imaging  
of sites expressing (II). (I) and (II) are useful for treating disorders  
involving aberrant protein expression or biological activity. The  
polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits to assess biodiversity  
and to produce other types of data and products dependent on DNA and  
amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
amino acid sequences of the invention. Note: The sequence data for this  
patent did not appear in the printed specification, but was obtained in  
electronic format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 140 AA;

Alignment Scores:  
Pred. No.: 3,038-67 Length: 140  
Score: 76.00 Matches: 76  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 24.76% Indels: 0  
DB: 4 Gaps: 0

AF334735 (1-954) x ABG11758 (1-140)

QY 310 GAATGGGGAGTAAGTAGAAGACCGCTTCTATACAATCATGCTCGAGGAGCAAGAA 369  
|||  
Db 65 GluTrpGlySerLysValGluAspArgPheTyrAsnAsnHisAlaPheGluGluGlu 84

QY 370 CCACCTGAGAAAGTGATCCTTAAACAAGAAGTCTCAGATATCTGGGAGGAGGAGAG 429  
|||  
Db 85 ProProGluLysSerAspProLysGlnGluSerGlnHisSerGlyLysGluGlu 104

QY 430 ACATCAGTCACCATCTTAGACTCTTCTGAGGAAGATAAGGAAAAAGAGAGGTTGCTGCT 489  
|||  
Db 105 ThrSerValThrIleLeuAspSerSerGluAspLysGluLysGluGluValAlaAla 124

QY 490 GTCAAAATCCAGCTGCTTCGGGGACACATAGCCAGAGAGGAGGCA 537  
|||  
Db 125 ValLysIleGlnAlaAlaPheArgGlyHisIleAlaAArgGluGluAla 140

RESULT 6  
ABG08152  
ID ABG08152 standard; protein; 300 AA.  
XX AC ABG08152;  
XX DT 13-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #8143.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US008631.  
XX PR 31-MAR-2000; 2000US-00540217.  
XX PR 23-AUG-2000; 2000US-00649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.

XX SQ Sequence 119 AA;  
Alignment Scores:  
Pred. No.: 3,098-67 Length: 119  
Score: 76.00 Matches: 76  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 24.76% Indels: 0  
DB: 3 Gaps: 0

AF334735 (1-954) x AAG00833 (1-119)

QY 136 ATGTCGATTCATTCTCCACACCCACTACCGAATCCACAGGATTTCGGATCTTCT 195  
|||  
Db 1 MetSerIleProPheSerAsnThrHisTyrArgIleProGlnGlyPheGlyAsnLeu 20

QY 196 GAAGGCTGACACGCGAGATTCTGAGAGAGCAACCGCAATATACACGCTTTTCGACGA 255  
|||  
Db 21 GluGlyLeuThrArgGluLeuArgGluGlnProAspAsnIleProAlaPheAlaAla 40

QY 256 GCCTATTTGAGAGCCTTCTAGAGAAAGAGAGAAACCAACTTTGATCCAGCAGATGG 315  
|||  
Db 41 AlaTyrPheGluSerLeuLeuGluLysArgGluLysThrAsnPheAspProAlaGluTrp 60

QY 316 GGGAGTAAGTAGAAGACCGCTTCTATACAATCATGCTTCGAGGAG 363  
|||  
Db 61 GlySerLysValGluAspArgPheTyrAsnAsnHisAlaPheGluGlu 76

RESULT 5  
ABG11758  
ID ABG11758 standard; protein; 140 AA.  
XX AC ABG11758;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #11749.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US008631.  
XX PR 31-MAR-2000; 2000US-00540217.  
XX PR 23-AUG-2000; 2000US-00649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX N-PSDB; AAS75945.

XX New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits and to assess  
biodiversity.

XX Claim 20; SEQ ID NO 42117; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
sequences. (I) is useful as hybridisation probes, polymerase chain  
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
and in recombinant production of (II). The polynucleotides are also used  
in diagnostics as expressed sequence tags for identifying expressed  
genes. (I) is useful in gene therapy techniques to restore normal

DR N-PSDB; AAS72339.  
 XX  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 XX Claim 20; SEQ ID NO 38511; 103pp; English.  
 PS  
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics; forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 300 AA;

Alignment Scores:  
 Pred. No.: 4,278-48 Length: 300  
 Score: 57.00 Matches: 57  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 18.57% Indels: 0  
 DB: 4 Gaps: 0

AF334735 (1-954) x ABG08152 (1-300)

QY 118 GCAGTTCTTACCAAGAGATGTCGATTCCATTCTCCACACCCACTACCGAATTCACCA 177  
 |||||  
 DB 223 AlaValLeuThrIlyslsMetSerIleProPheSerAsnThrHisTyArgIleProGln 242  
 QY 178 GGAATTTGGGAATCTTTGAGGGCTGACACGCGAGATCTGAGAGACGACCGGCAAT 237  
 |||||  
 DB 243 GlyPheGlyAsnLeuLeuGluGlyLeuThrArgGluIleLeuArgGluGlnProAspAsn 262  
 QY 238 ATACGAGCTTTTCAGCGCTATTTTGAGGCTTCTAGAGAAAGAGAG 288  
 |||||  
 DB 263 IleProAlaPheAlaAlaIlePheGluSerLeuLeuGluIlyslsArgGlu 279

RESULT 7

AAW42693  
 ID AAW42693 standard; protein; 163 AA.

AC AAW42693;

DT 27-APR-1998 (first entry)

DE Baboon sperm zona binding protein Sp17 (BSp17).

XX Sperm zona binding protein Sp17; zona pellucida; binding; reduction;  
 KW Human sperm autoantigen; antigen; antigenic epitope; immunocontraception;  
 KW fertility; immunocontraceptive vaccine; autoimmune fertility.

XX Papio sp.

OS

XX WO9739020-A2.

XX

PD 23-OCT-1997.  
 XX  
 XX 11-APR-1997; 97WO-US006489.  
 XX  
 XX 15-APR-1996; 96US-00632535.  
 XX  
 XX (UYNC-) UNIV NORTH CAROLINA.  
 PA  
 XX  
 XX Orand MG, Lea I, Widgren EE;  
 PI  
 XX  
 XX WPI; 1997-526394/48.  
 DR N-PSDB; AAV04821.  
 XX

PT Sperm protein Sp17 antigenic peptide(s) - used as immunocontraceptive to  
 reduce fertility.

XX Example 11; Fig 16A; 160pp; English.

CC The present sequence represents a baboon sperm zona binding protein Sp17.  
 CC The cDNA sequence was identified by screening a baboon testis cDNA  
 CC library with the coding region of the human Sp17 gene as a probe. The  
 CC human Sp17 protein binds to zona pellucida at high affinity by binding  
 CC sulphated, complex carbohydrates. Peptides comprising a contiguous  
 CC segment of the amino acids of human Sp17 have been found to be antigenic  
 CC in several species. Peptides AAW42774-867 contain antigenic epitopes in  
 CC mice, rabbits, non human primates and humans. The peptides can be used in  
 CC immunocontraceptive methods to reduce the fertility of animals,  
 CC particularly in females. The peptides can be used in an  
 CC immunocontraceptive vaccine. The presence of the antigenic peptides can  
 CC be detected to diagnose autoimmune fertility in both male and female  
 CC subjects

SQ Sequence 163 AA;

Alignment Scores:  
 Pred. No.: 6,388-34 Length: 163  
 Score: 43.00 Matches: 163  
 Percent Similarity: 95.89% Conservative: 0  
 Best Local Similarity: 95.89% Mismatches: 0  
 Query Match: 14.01% Indels: 3  
 DB: 2 Gaps: 0

AF334735 (1-954) x AAW42693 (1-163)

QY 136 ATGTCGATTCCATTCTCCACACCCACTACCGAATTCACCAAGGATTTCGGAATCTTCT 195  
 |||||  
 DB 1 MetSerIleProPheSerAsnThrHisTyArgIleProGlnGlyPheGlyAsnLeu 20  
 QY 196 GAAGGGCTGACACGCGAGATCTGAGAGACGACCGACAATATATACCACTTTTCAGCA 255  
 |||||  
 DB 21 GluGlyLeuThrArgGluIleLeuArgGluGlnProAspAsnIleProAlaPhe---Ala 39  
 QY 256 GCCTATTTGAGAGCTTCTAGAGAAAAGAGAGAAAACCACTTTGATCCAGCAGATGG 315  
 |||||  
 DB 40 AlaTyPheGluSerLeuLeuGluIlyslsArgGluIlyslsThrAsnPheAspProAlaGluTrp 59  
 QY 316 GCGAGTAGTAGAGACCGCTTCTATACAAAT 348  
 |||||  
 DB 60 GlySerLysValGluAspArgPheTyPheAsnAsn 70

RESULT 8

AAW42691  
 ID AAW42691 standard; protein; 146 AA.

AC AAW42691;

DT 27-APR-1998 (first entry)

DE Recombinant rabbit Sp17 protein.

XX Sp17; rabbit sperm antigen; RSA; sperm autoantigen; antigen;

KW antigenic epitope; immunocontraception; fertility; autoimmune fertility;  
 KW immunocontraceptive vaccine.



```

XX Synthetic.
OS Oryctolagus cuniculus.
XX
XX WO9739020-A2.
XX
XX 23-OCT-1997.
XX
XX 11-APR-1997; 97WO-US006489.
XX
XX 15-APR-1996; 96US-00632535.
XX
XX (UYNC-) UNIV NORTH CAROLINA.
XX
XX Orand MG, Lea I, Widgren EE;
XX
XX WPI; 1997-526394/48.
XX
XX Sperm protein Sp17 antigenic peptide(s) - used as immuno:contraceptive to
XX reduce fertility.
XX
XX Example 6; Page 79; 160pp; English.
XX
XX The present protein represents rabbit recombinant Sp17 protein. The
XX recombinant rabbit Sp17 protein is expressed minus the first 11 N-
XX terminal amino acids, but with an N-terminal containing the sequence Arg-
XX Gly-Ser, followed by 6 His and Gly-Ser, all of which precede the Sp17
XX amino acids. The binding of of immune sera from a female rabbit immunised
XX with the recombinant fusion Sp17 protein to the Sp17 sequential peptides
XX AAW42643-89 was determined. Peptides comprising a contiguous segment of
XX the amino acid of Sp17 have been found to be antigenic. They induce
XX antibodies which recognise sperm, and inhibit fertilisation. The peptides
XX can be used in immunocontraceptive methods as a immunocontraceptive
XX vaccine to reduce the fertility of animals, particularly in females. The
XX presence of the antigenic peptides can be detected to diagnose autoimmune
XX fertility in both male and female subjects
XX
XX Sequence 146 AA;
XX
XX Alignment Scores:
XX Pred. No.: 8.15e-24 Length: 146
XX Score: 33.00 Matches: 33
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 10.75% Indels: 0
XX DB: 2 Gaps: 0
XX
AF334735 (1-954) x AAW42691 (1-146)
XX
XX 169 ATTCCACAAGGATTTGGGAATCTTTGAAGGCTGACACGGAGATTCGAGAGACAA 228
XX
XX 12 IleProGlnGlyPheGlyAsnLeuLeuGluGlyLeuThrArgGluIleLeuArgGluIn 31
XX
XX 229 CCGACAATATACCACTTTTCAGCAGCAGCTATTGTGAG 267
XX
XX 32 ProAspAsnIleProAlaPheAlaAlaAlaTyPheGlu 44
XX
XX RESULT 9
XX AAW42767
XX ID AAW42767 standard; peptide; 28 AA.
XX
XX AAW42767;
XX
XX 27-APR-1998 (first entry)
XX
XX Peptide of the specification.
XX
XX Sp17; sperm autoantigen; antigen; antigenic epitope; fertility;
XX immunocontraception; immunocontraceptive vaccine; autoimmune fertility.
XX
XX Synthetic.
XX
XX WO9739020-A2.
XX

```

```

XX 23-OCT-1997.
XX
XX 11-APR-1997; 97WO-US006489.
XX
XX 15-APR-1996; 96US-00632535.
XX
XX (UYNC-) UNIV NORTH CAROLINA.
XX
XX Orand MG, Lea I, Widgren EE;
XX
XX WPI; 1997-526394/48.
XX
XX Sperm protein Sp17 antigenic peptide(s) - used as immuno:contraceptive to
XX reduce fertility.
XX
XX Disclosure; Page 77; 160pp; English.
XX
XX Peptides AAW42762-70 appear in the specification. A series of peptides
XX derived from rabbit and human Sp17 proteins are also disclosed. These
XX Sp17 peptides contain antigenic epitopes in mice, rabbits, non-human
XX primates and humans. They induce antibodies which recognise sperm, and
XX inhibit fertilisation. The peptides can be used in immunocontraceptive
XX methods as a immunocontraceptive vaccine to reduce the fertility of
XX animals, particularly in females. The presence of the antigenic peptides
XX can be detected to diagnose autoimmune fertility in both male and female
XX subjects
XX
XX Sequence 28 AA;
XX
XX Alignment Scores:
XX Pred. No.: 1.13e-18 Length: 28
XX Score: 28.00 Matches: 28
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 9.12% Indels: 0
XX DB: 2 Gaps: 0
XX
AF334735 (1-954) x AAW42767 (1-28)
XX
XX 298 TTGTATCCAGCAGCAATGGGAGTAGGTAGAGAGCCGCTTCTATACCAATCATCATTCC 357
XX
XX 1 PheAspProAlaGluTrpGlySerTyrValGluAspArgPheTyrAsnAsnHisAlaPhe 20
XX
XX 358 GAGGAGCAAGAACCACTTGAGAAA 381
XX
XX 21 GluGluGlnGluProProGluLys 28
XX
XX RESULT 10
XX AAW42762
XX ID AAW42762 standard; peptide; 25 AA.
XX
XX AAW42762;
XX
XX 27-APR-1998 (first entry)
XX
XX Peptide of the specification.
XX
XX Sp17; sperm autoantigen; antigen; antigenic epitope; fertility;
XX immunocontraception; immunocontraceptive vaccine; autoimmune fertility.
XX
XX Synthetic.
XX
XX WO9739020-A2.
XX
XX 23-OCT-1997.
XX
XX 11-APR-1997; 97WO-US006489.
XX
XX 15-APR-1996; 96US-00632535.
XX
XX (UYNC-) UNIV NORTH CAROLINA.
XX

```

PI Orand MG, Lea I, Widgren EB;  
 XX WPI; 1997-526394/48.  
 XX Sperm protein Sp17 antigenic peptide(s) - used as immuno:contraceptive to  
 PT reduce fertility.  
 PT Disclosure; Page 75; 160pp; English.  
 XX Peptides AAW42762-70 appear in the specification. A series of peptides  
 CC derived from rabbit and human Sp17 proteins are also disclosed. These  
 CC Sp17 peptides contain antigenic epitopes in mice, rabbits, non-human  
 CC primates and humans. They induce antibodies which recognise sperm, and  
 CC inhibit fertilisation. The peptides can be used in immunocontraceptive  
 CC methods as a immunocontraceptive vaccine to reduce the fertility of  
 CC animals, particularly in females. The presence of the antigenic peptides  
 CC can be detected to diagnose autoimmune fertility in both male and female  
 CC subjects  
 XX SQ Sequence 25 AA;  
 Alignment Scores:  
 Pred. No.: 1-22e-15 Length: 25  
 Score: 25.00 Matches: 25  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 8.14% Indels: 0  
 DB: 2 Gaps: 0  
 AF334735 (1-954) x AAW42762 (1-25)  
 QY 145 CCATTCACCAACCCACTACCGAATTCACAGGATTTGGGAATCTTCTCAAGGCTG 204  
 Db 1 ProPheSerAsnThrHisTyArgIleProGlnGlyPheGlyAsnLeuGluGlyLeu 20  
 QY 205 ACACGGCAGATTCTG 219  
 Db 21 ThrArgGluIleLeu 25  
 RESULT 11  
 ABG11759  
 ID ABG11759 standard; protein; 273 AA.  
 AC ABG11759;  
 XX 18-FEB-2002 (first entry)  
 DT Novel human diagnostic protein #11750.  
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 OS WO200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US008631.  
 XX 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX (HYSB-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 PI WPI; 2001-639362/73.  
 DR N-PSDB; AAS75946.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess

PT biodiversity.  
 XX Claim 20; SEQ ID NO 42118; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIFO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 273 AA;  
 Alignment Scores:  
 Pred. No.: 9.07e-16 Length: 273  
 Score: 25.00 Matches: 25  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 8.14% Indels: 0  
 DB: 4 Gaps: 0  
 AF334735 (1-954) x ABG11759 (1-273)  
 QY 436 GTCACATCTTAGACTCTTCTGAGGAAGATAGGAAGAAAGAGAGTCTGTCGCAA 495  
 Db 62 ValThrIleLeuAspSerGluGluAspIysGluLysGluLysValAlaValLys 81  
 QY 496 ATCCAAGCTGCCTTC 510  
 Db 82 IleGlnAlaAlaPhe 86  
 RESULT 12  
 AAW42771  
 ID AAW42771 standard; peptide; 21 AA.  
 AC AAW42771;  
 XX 27-APR-1998 (first entry)  
 DT Peptide derived from Sp17.  
 DE Sp17; sperm autoantigen; antigen; antigenic epitope; fertility;  
 KW immunocontraception; immunocontraceptive vaccine; autoimmune fertility.  
 XX Synthetic.  
 OS WO9739020-A2.  
 PN 23-OCT-1997.  
 PD 11-APR-1997; 97WO-US006489.  
 PF 15-APR-1996; 96US-00632535.  
 PR (UYNC-) UNIV NORTH CAROLINA.  
 PA Orand MG, Lea I, Widgren EB;  
 XX

```
DR WPI; 1997-526394/48.
XX
PT Sperm protein Sp17 antigenic peptide(s) - used as immuno:contraceptive to
PT reduce fertility.
XX
PS Disclosure; Page 83; 160pp; English.
XX
CC The present peptide is derived from a Sp17 protein. Sp17 peptides contain
CC antigenic epitopes in mice, rabbits, non-human primates and humans. They
CC induce antibodies which recognise sperm, and inhibit fertilisation. The
CC peptides can be used in immunocontraceptive methods as a
CC immunocontraceptive vaccine to reduce the fertility of animals,
CC particularly in females. The presence of the antigenic peptides can be
CC detected to diagnose autoimmune fertility in both male and female
CC subjects
XX
SQ Sequence 21 AA;
Alignment Scores:
Pred. No.: 1.37e-11 Length: 21
Score: 21.00 Matches: 21
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.84% Indels: 0
DB: 2 Gaps: 0
AF334735 (1-954) x AAW42771 (1-21)
QY 166 CGAATTCACAGAGATTTCGGATCTTGAAGGCTGACACGCGAGATTCTGAGAGAG 225
Db |||||
1 ArgileProGlnGlyPheGlyAsnLeuLeuGluGlyLeuThrArgGluIleLeuArgGlu 20
QY 226 CAA 228
Db |||
21 Gln 21
RESULT 13
AAW42770
ID AAW42770 standard; peptide; 21 AA.
XX
AC AAW42770;
XX
DT 27-APR-1998 (first entry)
XX
DE Peptide of the specification.
XX
KW Sp17; sperm autoantigen; antigen; antigenic epitope; fertility;
KW immunocontraception; immunocontraceptive vaccine; autoimmune fertility.
XX
OS Synthetic.
XX
PN WO9739020-A2.
XX
PD 23-OCT-1997.
XX
PF 11-APR-1997; 97WO-US006489.
XX
PR 15-APR-1996; 96US-00632535.
XX
PA (UYNC-) UNIV NORTH CAROLINA.
XX
PI Orand MG, Lea I, Widgren EE;
XX
DR WPI; 1997-526394/48.
XX
PT Sperm protein Sp17 antigenic peptide(s) - used as immuno:contraceptive to
PT reduce fertility.
XX
PS Disclosure; Page 78; 160pp; English.
XX
CC Peptides AAW42762-70 appear in the specification. A series of peptides
CC derived from rabbit and human Sp17 proteins are also disclosed. These
CC Sp17 peptides contain antigenic epitopes in mice, rabbits, non-human
CC
```

```
CC primates and humans. They induce antibodies which recognise sperm, and
CC inhibit fertilisation. The peptides can be used in immunocontraceptive
CC methods as a immunocontraceptive vaccine to reduce the fertility of
CC animals, particularly in females. The presence of the antigenic peptides
CC can be detected to diagnose autoimmune fertility in both male and female
CC subjects
XX
SQ Sequence 21 AA;
Alignment Scores:
Pred. No.: 1.37e-11 Length: 21
Score: 21.00 Matches: 21
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.84% Indels: 0
DB: 2 Gaps: 0
AF334735 (1-954) x AAW42770 (1-21)
QY 481 GTTGCTGCTGTCAAAATCCAAAGCTGCCCTTCGGGACACATAGCCAGAGAGGAGCAAG 540
Db |||||
1 ValAlaAlaVallylleGlnAlaAlaPheArgGlyHisIleAlaArgGluAlaLys 20
QY 541 AAA 543
Db |||
21 Lys 21
RESULT 14
AAW42772
ID AAW42772 standard; peptide; 21 AA.
XX
AC AAW42772;
XX
DT 27-APR-1998 (first entry)
XX
DE Peptide derived from Sp17.
XX
KW Sp17; sperm autoantigen; antigen; antigenic epitope; fertility;
KW immunocontraception; immunocontraceptive vaccine; autoimmune fertility.
XX
OS Synthetic.
XX
PN WO9739020-A2.
XX
PD 23-OCT-1997.
XX
PF 11-APR-1997; 97WO-US006489.
XX
PR 15-APR-1996; 96US-00632535.
XX
PA (UYNC-) UNIV NORTH CAROLINA.
XX
PI Orand MG, Lea I, Widgren EE;
XX
DR WPI; 1997-526394/48.
XX
PT Sperm protein Sp17 antigenic peptide(s) - used as immuno:contraceptive to
PT reduce fertility.
XX
PS Claim 1; Page 83; 160pp; English.
XX
CC The present peptide is derived from a Sp17 protein. Sp17 peptides contain
CC antigenic epitopes in mice, rabbits, non-human primates and humans. They
CC induce antibodies which recognise sperm, and inhibit fertilisation. The
CC peptides can be used in immunocontraceptive methods as a
CC immunocontraceptive vaccine to reduce the fertility of animals,
CC particularly in females. The presence of the antigenic peptides can be
CC detected to diagnose autoimmune fertility in both male and female
CC subjects
XX
SQ Sequence 21 AA;
Alignment Scores:
```

Pred. No.: 1.37e-11 Length: 21  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 6.84% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x AAW42772 (1-21)

QY 319 AGTAAGGTAGACGCTTCTATACATCATGCGAGGAGCAAGAACCCCTGAG 378  
Db 1 SerLysValGluAspArgPheTyAsnAsnHisAlaPheGluGluProGlu 20

QY 379 AAA 381  
Db 21 Lys 21

RESULT 15

AAW42869

ID AAW42869 standard; peptide; 19 AA.

XX AC AAW42869;

XX DT 27-APR-1998 (first entry)

XX DE Peptide of the specification.

XX KW Sp17; sperm autoantigen; antigen; antigenic epitope; fertility;

XX KM immunocontraception; immunocontraceptive vaccine; autoimmune fertility.

XX OS Synthetic.

XX PN WO9739020-A2.

XX PD 23-OCT-1997.

XX PF 11-APR-1997; 97WO-US006489.

XX PR 15-APR-1996; 96US-00632535.

XX PA (UYNC-) UNIV NORTH CAROLINA.

XX PI Orand MG, Lea I, Widgren BE;

XX PS WPI; 1997-526394/48.

XX PT Sperm protein Sp17 antigenic peptide(s) - used as immuno:contraceptive to

XX PT reduce fertility.

XX PS Disclosure; Page 134; 160pp; English.

XX CC The present peptide appears in the specification. A series of peptides  
XX CC derived from rabbit and human Sp17 proteins are also disclosed. These  
XX CC Sp17 peptides contain antigenic epitopes in mice, rabbits, non-human  
XX CC primates and humans. They induce antibodies which recognise sperm, and  
XX CC inhibit fertilisation. The peptides can be used in immunocontraceptive  
XX CC methods as a immunocontraceptive vaccine to reduce the fertility of  
XX CC animals, particularly in females. The presence of the antigenic peptides  
XX CC can be detected to diagnose autoimmune fertility in both male and female  
XX CC subjects

SQ Sequence 19 AA;

Alignment Scores:

Pred. No.: 1.46e-09 Length: 19  
Score: 19.00 Matches: 19  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 6.19% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x AAW42869 (1-19)

QY 487 GCTGTCAAATCCAAAGTCCTTCGGGGACACATAGCCAGAGAGGCAAGAAA 543  
Db 1 AlaValLysIleGlnAlaAlaPheArgGlyHisIleAlaArgGluGluAlaLysLys 19

Search completed: February 15, 2005, 02:22:02  
Job time : 171 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 15, 2005, 02:14:06 ; Search time 37.5 Seconds  
(without alignments)  
3798.144 Million cell updates/sec

Title: AF334735  
Perfect score: 307  
Sequence: 1 TCGCCCTCTCGCGCGCG.....AAAAAAAAAAAAAAAAAAAA 954

Scoring table:

OLIGO  
Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 513545 seqs, 74649064 residues

Word size: 1

Total number of hits satisfying chosen parameters: 903960

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

MODEL=frame+ n2p.model -DEV=xlp  
-Q/cgn2\_1/USPTO\_spool\_P/AF334735/runat\_14022005\_160239\_19363/app\_query.fasta\_1.1095  
-DB=issued\_Patents\_AA -OPT=fastan -SUFFIX=oligo.ra1 -MINMATCH=0.1 -LOOPCL=70  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=AF334735 -CGN 1 1 46 @runat\_14022005\_160239\_19363 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_AA:  
1: /cgn2\_6/prodata/1/iaa/5A COMB.pep:\*  
2: /cgn2\_6/prodata/1/iaa/5B COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/6A COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/6B COMB.pep:\*  
5: /cgn2\_6/prodata/1/iaa/PCTUS COMB.pep:\*  
6: /cgn2\_6/prodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	151	49.2	151	1	Sequence 2, Appl1
2	151	49.2	151	2	Sequence 2, Appl1
3	151	49.2	151	4	Sequence 2, Appl1
4	76	24.8	119	4	Sequence 4914, Ap
5	44	14.3	145	1	Sequence 50, Appl
6	44	14.3	146	2	Sequence 50, Appl
7	44	14.3	146	2	Sequence 50, Appl
8	44	14.3	148	1	Sequence 51, Appl
9	44	14.3	148	2	Sequence 51, Appl
10	44	14.3	148	2	Sequence 51, Appl
11	28	9.1	28	2	Sequence 57, Appl
12	28	9.1	28	2	Sequence 57, Appl

ALIGNMENTS

RESULT 1  
US-08-166-195A-2  
; Sequence 2, Application US/08166195A  
; Patent No. 5480799  
; GENERAL INFORMATION:  
; APPLICANT: O'Rand, Michael G.  
; APPLICANT: Widgren, Esther E.  
; APPLICANT: Richardson, Richard T.  
; APPLICANT: Lea, Isabel  
; TITLE OF INVENTION: Sperm Antigen Corresponding to a  
; TITLE OF INVENTION: Sperm Zona Binding Protein Autoantigenic Epitope  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenneth D. Sibley  
; STREET: P.O. Box 34009  
; CITY: Charlotte  
; STATE: No. 5480799th Carolina  
; COUNTRY: USA  
; ZIP: 28234  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/166,195A  
; FILING DATE: 10 DEC 1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sibley, Kenneth D.  
; REGISTRATION NUMBER: 31,665  
; REFERENCE/DOCKET NUMBER: 5470/73  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-881-3140  
; TELEFAX: 919-881-3175  
; INFORMATION FOR SEQ ID NO: 2:

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-166-195A-2

Alignment Scores:
Pred. No.: 2,346-135 Length: 151
Score: 151.00 Matches: 151
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 49.19% Indels: 0
DB: Gaps: 0

AF334735 (1-954) x US-08-166-195A-2 (1-151)

QY 136 ATGTCGATTCATTCCTCCACACCCACTACCGAATTCACAGGATTTGGGAATCTTCTT 195
Db 1 MetSerIleProPheSerAsnThrHisTyrArgIleProGlnGlyPheGlyAsnLeuLeu 20
QY 196 GAAGGCTGACGCGAGATTCCTGAGAGACCAACCGGACAATATACCACTTTTGCAGCA 255
Db 21 GluGlyLeuThrArgGluIleLeuArgGlnProAspAsnIleProAlaPheAlaAla 40
QY 256 GCCTATTGAGAGCCCTCTAGAGAAAGAGGAAACCAACTTTGATCCAGCAGATGG 315
Db 41 AlaTyrPheGluSerLeuLeuGluLysArgGluLysThrAsnPheAspProAlaGluTrp 60
QY 316 GGGAGTAAGTAGAAGACCGCTTCTATACCAATCATGCAATTCGAGGAGCAAGACCACT 375
Db 61 GlySerLysValGluAspArgPheTyrAsnAsnHisAlaPheGluGlnGluProPro 80
QY 376 GAGAAAAGTGTCTTAAACAAGAGAGTCTCAGATATCTGGGAGGAGGAGGAGGAGCATCA 435
Db 81 GluLysSerAspProLysGlnGluSerGlnIleSerGlyLysGluGluThrSer 100
QY 436 GTCACCATCTTAGACTCTTCTGAGGAAGATAGGAAAAGAGAGGTTTGGCTGCTCAAA 495
Db 101 ValThrIleLeuAspSerSerGluGluAspLysGluLysGluGluValAlaAlaValLys 120
QY 496 ATCCAAGCTGCCTTCGGGGACACATAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 555
Db 121 IleGlnAlaAlaPheArgGlyHisIleAlaArgGluGluAlaLysMetLysThrAsn 140
QY 556 AGTCTTCAAAATGAGGAAAAAGGAAAAACAAG 588
Db 141 SerLeuGlnAsnGluLysGluGluAsnLys 151

RESULT 3
US-08-436-772-2
; Sequence 2, Application US/08436772
; Patent No. 5814456
; GENERAL INFORMATION:
; APPLICANT: O'Rand, Michael G.
; APPLICANT: Widgren, Esther E.
; APPLICANT: Richardson, Richard T.
; APPLICANT: Lea, Isabel
; TITLE OF INVENTION: Sperm Antigen Corresponding to a Sperm
; TITLE OF INVENTION: Zona Binding Protein Autoantigenic Epitope
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley
; STREET: P.O. Box 34009
; CITY: Charlotte
; STATE: No. 5814456th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,772
; FILING DATE: 08-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5470-73B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-881-3140
; TELEFAX: 919-881-3175
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-436-772-2

Alignment Scores:
Pred. No.: 2,346-135 Length: 151
Score: 151.00 Matches: 151
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 49.19% Indels: 0
DB: Gaps: 0

AF334735 (1-954) x US-08-436-772-2 (1-151)

QY 136 ATGTCGATTCATTCCTCCACACCCACTACCGAATTCACAGGATTTGGGAATCTTCTT 195
Db 1 MetSerIleProPheSerAsnThrHisTyrArgIleProGlnGlyPheGlyAsnLeuLeu 20
QY 196 GAAGGCTGACGCGAGATTCCTGAGAGACCAACCGGACAATATACCACTTTTGCAGCA 255
Db 21 GluGlyLeuThrArgGluIleLeuArgGlnProAspAsnIleProAlaPheAlaAla 40
QY 256 GCCTATTGAGAGCCCTCTAGAGAAAGAGGAAACCAACTTTGATCCAGCAGATGG 315
Db 41 AlaTyrPheGluSerLeuLeuGluLysArgGluLysThrAsnPheAspProAlaGluTrp 60
QY 316 GGGAGTAAGTAGAAGACCGCTTCTATACCAATCATGCAATTCGAGGAGCAAGACCACT 375
Db 61 GlySerLysValGluAspArgPheTyrAsnAsnHisAlaPheGluGlnGluProPro 80
QY 376 GAGAAAAGTGTCTTAAACAAGAGAGTCTCAGATATCTGGGAGGAGGAGGAGGAGCATCA 435
Db 81 GluLysSerAspProLysGlnGluSerGlnIleSerGlyLysGluGluThrSer 100
QY 436 GTCACCATCTTAGACTCTTCTGAGGAAGATAGGAAAAGAGAGGTTTGGCTGCTCAAA 495
Db 101 ValThrIleLeuAspSerSerGluGluAspLysGluLysGluGluValAlaAlaValLys 120
QY 496 ATCCAAGCTGCCTTCGGGGACACATAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 555
Db 121 IleGlnAlaAlaPheArgGlyHisIleAlaArgGluGluAlaLysMetLysThrAsn 140
QY 556 AGTCTTCAAAATGAGGAAAAAGGAAAAACAAG 588
Db 141 SerLeuGlnAsnGluLysGluGluAsnLys 151

RESULT 3
US-08-436-883B-2
; Sequence 2, Application US/08436883B
; Patent No. 5820861
; GENERAL INFORMATION:
; APPLICANT: O'Rand, Michael G.
; APPLICANT: Widgren, Esther E.
; APPLICANT: Richardson, Richard T.
; APPLICANT: Lea, Isabel
; TITLE OF INVENTION: Sperm Antigen Corresponding to a Sperm
; TITLE OF INVENTION: Zona Binding Protein Autoantigenic Epitope
; NUMBER OF SEQUENCES: 60
```



;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Kenneth D. Sibley  
;; STREET: P.O. Box 34009  
;; CITY: Charlotte  
;; STATE: No. 5820861th Carolina  
;; COUNTRY: USA  
;; ZIP: 28234  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/436,883B  
;; FILING DATE: 08-MAY-1995  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Sibley, Kenneth D.  
;; REGISTRATION NUMBER: 31,665  
;; REFERENCE/DOCKET NUMBER: 5470-73C  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 919-420-2200  
;; TELEFAX: 919-881-3175  
;;  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 151 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-436-883B-2

Alignment Scores:  
Pred. No.: 2,34e-135 Length: 151  
Score: 151.00 Matches: 151  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 49.19% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x US-08-436-883B-2 (1-151)

```
QY 136 ATGTCGATTCATTCTCCACACCCACTACCGAATCCACAGATTTGGGAATCTTCTT 195
Db 1 MetSerIleProPheSerAsnThrHisTyrArgIleProGlnGlyPheGlyAsnLeu 20
QY 196 GAGGGCTGACGCGGAGATTCTCAGAGACCAACCGACATATACGAGTTTTCAGCA 255
Db 21 GluGlyLeuThrArgGluIleLeuArgGluGlnProAspAsnIleProAlaPheAla 40
QY 256 GCCTATTTCAGAGCCTTCTAGAGAAAGAGAGAAACCAACTTTCATCCAGCAGAAATGG 315
Db 41 AlaTyrPheGluSerLeuLeuGluLysArgGluLysThrAsnPheAspProAlaGluT 60
QY 316 GGGAGTAAGTAGAAGCCGCTTCTATAACAATCATGCAATTCGAGGAGCAAGAACCACT 375
Db 61 GlySerIysValGluAspArgPheTyrAsnAsnHisAlaPheGluGlnGluProPro 80
QY 376 GAGAAAGTCACTCTAAACAAGAGCTCTCAGATATCTCGGAAGGAGGAGAGACATCA 435
Db 81 GluLysSerAspProLysGlnGluSerGlnIleSerGlyLysGluGluGluThrSer 100
QY 436 GTCAACATCTAGACTCTCTGAGGAGAGATAAGGAAAGAAAGAGAGGTTGCTGTGCAAA 495
Db 101 ValThrIleLeuAspSerSerGluGluAspLysGluValAlaIaValLys 120
QY 496 ATCCAACTGCCTTCCGGGACACATAGCCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 555
Db 121 IleGlnAlaAlaPheArgGlyHisIleAlaArgGluGluAlaLysMetLysThrAsn 140
QY 556 AGTCTTCAAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 589
Db 141 SerLeuGlnAsnGluGluLysGluGluAsnLys 151
```

#### RESULT 4

US-09-513-999C-4914  
; Sequence 4914, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; Patent No. 6783961  
; FILE REFERENCE: 59 US2, REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 4914  
; LENGTH: 119  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: 77  
; OTHER INFORMATION: Xaa=His or Leu or Gln  
US-09-513-999C-4914

Alignment Scores:  
Pred. No.: 7,47e-64 Length: 119  
Score: 76.00 Matches: 76  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 24.76% Indels: 0  
DB: 4 Gaps: 0

AF334735 (1-954) x US-09-513-999C-4914 (1-119)

```
QY 136 ATGTCGATTCATTCTCCACACCCACTACCGAATTCACAGATTTGGGAATCTTCTT 195
Db 1 MetSerIleProPheSerAsnThrHisTyrArgIleProGlnGlyPheGlyAsnLeu 20
QY 196 GAGGGCTGACGCGGAGATTCTCAGAGACCAACCGACATATACGAGTTTTCAGCA 255
Db 21 GluGlyLeuThrArgGluIleLeuArgGluGlnProAspAsnIleProAlaPheAla 40
QY 256 GCCTATTTCAGAGCCTTCTAGAGAAAGAGAGAAACCAACTTTCATCCAGCAGAAATGG 315
Db 41 AlaTyrPheGluSerLeuLeuGluLysArgGluLysThrAsnPheAspProAlaGluT 60
QY 316 GGGAGTAAGTAGAAGCCGCTTCTATAACAATCATGCAATTCGAGGAG 363
Db 61 GlySerLysValGluAspArgPheTyrAsnAsnHisAlaPheGluGlu 76
```

#### RESULT 5

US-08-166-195A-50  
; Sequence 50, Application US/08166195A  
; Patent No. 5480799  
; GENERAL INFORMATION:

;; APPLICANT: O'Rand, Michael G.  
;; APPLICANT: Widgren, Esther E.  
;; APPLICANT: Richardson, Richard T.  
;; APPLICANT: Lea, Isabel  
;; TITLE OF INVENTION: Sperm Antigen Corresponding to a  
;; TITLE OF INVENTION: Sperm Zona Binding Protein Autoantigenic Epitope  
;; NUMBER OF SEQUENCES: 51  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Kenneth D. Sibley  
;; STREET: P.O. Box 34009  
;; CITY: Charlotte  
;; STATE: No. 5480799th Carolina  
;; COUNTRY: USA  
;; ZIP: 28234  
;; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/166,195A  
FILING DATE: 10 DEC 1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sibley, Kenneth D.  
REGISTRATION NUMBER: 31,665  
REFERENCE/DOCKET NUMBER: 5470/73  
TELEPHONE: 919-881-3140  
TELEFAX: 919-881-3175  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 145 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-166-195A-50

Alignment Scores:  
Pred. No.: 2,3e-33 Length: 145  
Score: 44.00 Matches: 44  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 14.33% Indels: 0  
DB: 1 Gaps: 0

AF334735 (1-954) x US-08-166-195A-50 (1-145)

QY 136 ATGTCGATTCATCTCCACACCCACTACCGAATTCACAGGATTTGGGAATCTTCTT 195  
Db 1 MetSerIleProPheSerAsnThrHisTyArgIleProGlnGlyPheGlyAsnLeu 20  
QY 196 GAAGGCTGACGCGAGATTCTGAGAGACGACGACCAATATACCGCTTTTCAGCA 255  
Db 21 GluGlyLeuThrArgGluLeuArgGluGlnProAspAsnIleProAlaPheAla 40

QY 256 GCCTATTTTGAG 267  
Db 41 AlaTyPheGlu 44

RESULT 6  
US-08-436-772-50  
Sequence 50, Application US/08436772  
Patent No. 5814456  
GENERAL INFORMATION:  
APPLICANT: O'Rand, Michael G.  
APPLICANT: Widgren, Esther E.  
APPLICANT: Richardson, Richard T.  
APPLICANT: Lea, Isabel  
TITLE OF INVENTION: Sperm Antigen Corresponding to a Sperm  
TITLE OF INVENTION: Zona Binding Protein Autoantigenic Epitope  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenneth D. Sibley  
STREET: P.O. Box 34009  
CITY: Charlotte  
STATE: No. 5814456th Carolina  
COUNTRY: USA  
ZIP: 28234  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/436,772  
FILING DATE: 08-MAY-1995

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sibley, Kenneth D.  
REGISTRATION NUMBER: 31,665  
REFERENCE/DOCKET NUMBER: 5470-73B  
TELEPHONE: 919-881-3140  
TELEFAX: 919-881-3175  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 146 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-436-772-50

Alignment Scores:  
Pred. No.: 2,29e-33 Length: 146  
Score: 44.00 Matches: 44  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 14.33% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x US-08-436-772-50 (1-146)

QY 136 ATGTCGATTCATCTCCACACCCACTACCGAATTCACAGGATTTGGGAATCTTCTT 195  
Db 1 MetSerIleProPheSerAsnThrHisTyArgIleProGlnGlyPheGlyAsnLeu 20  
QY 196 GAAGGCTGACGCGAGATTCTGAGAGACGACGACCAATATACCGCTTTTCAGCA 255  
Db 21 GluGlyLeuThrArgGluLeuArgGluGlnProAspAsnIleProAlaPheAla 40

QY 256 GCCTATTTTGAG 267  
Db 41 AlaTyPheGlu 44

RESULT 7  
US-08-436-883B-50  
Sequence 50, Application US/08436883B  
Patent No. 5820861  
GENERAL INFORMATION:  
APPLICANT: O'Rand, Michael G.  
APPLICANT: Widgren, Esther E.  
APPLICANT: Richardson, Richard T.  
APPLICANT: Lea, Isabel  
TITLE OF INVENTION: Sperm Antigen Corresponding to a Sperm  
TITLE OF INVENTION: Zona Binding Protein Autoantigenic Epitope  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenneth D. Sibley  
STREET: P.O. Box 34009  
CITY: Charlotte  
STATE: No. 5820861th Carolina  
COUNTRY: USA  
ZIP: 28234  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/436,883B  
FILING DATE: 08-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sibley, Kenneth D.  
REGISTRATION NUMBER: 31,665  
REFERENCE/DOCKET NUMBER: 5470-73C  
TELEPHONE: 919-420-2200

TELEFAX: 919-881-3175	MOLECULE TYPE: protein	Length: 148
INFORMATION FOR SEQ ID NO: 50:		
SEQUENCE CHARACTERISTICS:		
LENGTH: 146 amino acids	2,29e-33	Matches: 44
TYPE: amino acid	Score: 44.00	Conservative: 0
STRANDEDNESS: single	Percent Similarity: 100.00%	Mismatches: 0
TOPOLOGY: linear	Best Local Similarity: 100.00%	Indels: 0
MOLECULE TYPE: peptide	Query Match: 14.33%	Gaps: 0
US-08-436-883B-50		
Alignment Scores:		
Pred. No.: 2,29e-33		
Score: 44.00		
Percent Similarity: 100.00%		
Best Local Similarity: 100.00%		
Query Match: 14.33%		
DB: 2		
AP334735 (1-954) x US-08-436-883B-50 (1-146)		
QY 136 ATGTCGATTCATCTCCACACCCACTACCGAATTCACAGGATTGGGAATCTCTT 195		
Db 1 MetSerIleProPheSerAsnThrHisTyArgIleProGInglyPheGlyAsnLeu 20		
QY 196 GAAGGCTGACACGCGAGATCTGAGAGACGACCGACATATACCGCTTTTCGACGA 255		
Db 21 GluGlyLeuThrArgGluLeuArgGluGlnProAspAsnIleProAlaPheAla 40		
QY 256 GCCTATTTGAG 267		
Db 41 AlaTyPheGlu 44		
RESULT 8		
US-08-166-195A-51		
Sequence 51, Application US/08166195A		
Patent No. 5480793		
GENERAL INFORMATION:		
APPLICANT: O'Rand, Michael G.		
APPLICANT: Widgren, Esther E.		
APPLICANT: Richardson, Richard T.		
APPLICANT: Lea, Isabel		
TITLE OF INVENTION: Sperm Antigen Corresponding to a		
TITLE OF INVENTION: Sperm Zona Binding Protein Autoantigenic Epitope		
NUMBER OF SEQUENCES: 51		
CORRESPONDENCE ADDRESS:		
ADDRESSEE: Kenneth D. Sibley		
STREET: P.O. Box 34009		
CITY: Charlotte		
STATE: No. 5480799th Carolina		
COUNTRY: USA		
ZIP: 28234		
COMPUTER READABLE FORM:		
MEDIUM TYPE: Floppy disk		
OPERATING SYSTEM: PC-DOS/MS-DOS		
SOFTWARE: Patent In Release #1.0, Version #1.25		
CURRENT APPLICATION DATA:		
APPLICATION NUMBER: US/08/166, 195A		
FILING DATE: 10 DEC 1993		
CLASSIFICATION: 435		
ATTORNEY/AGENT INFORMATION:		
NAME: Sibley, Kenneth D.		
REGISTRATION NUMBER: 31,665		
REFERENCE/DOCKET NUMBER: 5470/73		
TELEPHONE: 919-881-3175		
TELEFAX: 919-881-3175		
INFORMATION FOR SEQ ID NO: 51:		
SEQUENCE CHARACTERISTICS:		
LENGTH: 148 amino acids		
TYPE: amino acid		
STRANDEDNESS: single		
TOPOLOGY: linear		
US-08-436-883B-50		
Alignment Scores:		
Pred. No.: 2,29e-33		
Score: 44.00		
Percent Similarity: 100.00%		
Best Local Similarity: 100.00%		
Query Match: 14.33%		
DB: 2		
AP334735 (1-954) x US-08-166-195A-51 (1-148)		
QY 136 ATGTCGATTCATCTCCACACCCACTACCGAATTCACAGGATTGGGAATCTCTT 195		
Db 1 MetSerIleProPheSerAsnThrHisTyArgIleProGInglyPheGlyAsnLeu 20		
QY 196 GAAGGCTGACACGCGAGATCTGAGAGACGACCGACATATACCGCTTTTCGACGA 255		
Db 21 GluGlyLeuThrArgGluLeuArgGluGlnProAspAsnIleProAlaPheAla 40		
QY 256 GCCTATTTGAG 267		
Db 41 AlaTyPheGlu 44		
RESULT 9		
US-08-436-772-51		
Sequence 51, Application US/08436772		
Patent No. 5814456		
GENERAL INFORMATION:		
APPLICANT: O'Rand, Michael G.		
APPLICANT: Widgren, Esther E.		
APPLICANT: Richardson, Richard T.		
APPLICANT: Lea, Isabel		
TITLE OF INVENTION: Sperm Antigen Corresponding to a Sperm		
TITLE OF INVENTION: Zona Binding Protein Autoantigenic Epitope		
NUMBER OF SEQUENCES: 60		
CORRESPONDENCE ADDRESS:		
ADDRESSEE: Kenneth D. Sibley		
STREET: P.O. Box 34009		
CITY: Charlotte		
STATE: No. 5814456th Carolina		
COUNTRY: USA		
ZIP: 28234		
COMPUTER READABLE FORM:		
MEDIUM TYPE: Floppy disk		
OPERATING SYSTEM: PC-DOS/MS-DOS		
SOFTWARE: Patent In Release #1.0, Version #1.30		
CURRENT APPLICATION DATA:		
APPLICATION NUMBER: US/08/436, 772		
FILING DATE: 08-MAY-1995		
CLASSIFICATION: 435		
ATTORNEY/AGENT INFORMATION:		
NAME: Sibley, Kenneth D.		
REGISTRATION NUMBER: 31,665		
REFERENCE/DOCKET NUMBER: 5470-73B		
TELEPHONE: 919-881-3140		
TELEFAX: 919-881-3175		
INFORMATION FOR SEQ ID NO: 51:		
SEQUENCE CHARACTERISTICS:		
LENGTH: 148 amino acids		
TYPE: amino acid		

```
Best Local Similarity: 100.00%
Query Match: 14.33%
DB: 2
Mismatches: 0
Indels: 0
Gaps: 0
```

AF334735 (1-954) x US-08-436-772-51 (1-148)

136	ATGTCGATTCCATTCTCCAAACACCCTACTCCAGGATTCCACAGGATTGGGAATCTCTT	195
QY		
1	MetSerIleProPheSerAsnThrHisTyrargIleProGlnGlyPheGlyAsnLeu	20
Db		
196	GAAGGGCTGACACGCGAGATTCTGAGAGACCAACCGCAATATACACAGCTTTTGCAGCA	255
QY		
21	GluGlyLeuThrArgGluIleLeuArgGluGlnProAspAsnIleProAlaPheAlaAla	40
Db		

QY	256	GCCTATT	267
Db	41	AlaTyrPheGlu	44

RESULT 10  
US-08-436--883B-51  
; Sequence 51, Application US/08436883B  
; Patent No. 5820861  
; GENERAL INFORMATION:  
; APPLICANT: O'Rand, Michael G.  
; APPLICANT: Widgren, Esther E.  
; APPLICANT: Richardson, Richard T.  
; APPLICANT: Lea, Isabel  
; TITLE OF INVENTION: Sperm Antigen Corresponding to a Sperm  
; TITLE OF INVENTION: Zona Binding Protein Autoantigenic Epitope

Alignment Scores:		
Pred. No.:	2,298-33	Length: 148
Score:	44.00	Matches: 44
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	14.33%	Indels: 0
DB:	2	Gaps: 0

AF334735 (1-954) x US-08-436-883B-51 (1-148)

QY 136 ATGTCGATTCCATTCTCCAACACCCACTACCGAATTCACAAGGATTGGGAATCTTCTT 195

Db 1 MetSerIleProPheSerAsnThrHisTyrArgIleProGlnGlyPheGlyAsnLeu 20  
 Qy 196 GAAGGGCTCACCGCGAGATTCTGAGAGAGCAACCGACAATATACCAAGCTTTTGAGCA 255  
 Db 21 GluGlyLeuThrArgGluIleLeuArgGluGlnProAspAsnIleProAlaPheAlaAla 40  
 Qy 256 GCCTATTATTGAG 267  
 Db 41 AlaTyrPheGlu 44

```

RESULT 11
US-08-436-772-57
; Sequence 57, Application US/08436772
; Patent No. 5814456
; GENERAL INFORMATION:
; APPLICANT: O'Rand, Michael G.
; APPLICANT: Widgren, Esther E.
; APPLICANT: Richardson, Richard T.
; APPLICANT: Lea, Isabel
; TITLE OF INVENTION: Sperm Antigen Corresponding to a Sperm
; TITLE OF INVENTION: Zona Binding Protein Autoantigenic Epitope
; NUMBER OF SEQUENCES: 60

```

Alignment Scores:	
Pred. No.:	5,31e-18
Score:	28.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	9.12%
DB:	2
Length:	28
Matches:	28
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

AF334735 (1-954) x US-08-436-772-57 (1-28)

[illegible]

## RESULT 12

US-08-436-883B-57  
; Sequence 57, Application US/08436883B  
; Patent No. 5820861  
; GENERAL INFORMATION:  
; APPLICANT: O'Rand, Michael G.  
; APPLICANT: Widgren, Esther E.  
; APPLICANT: Richardson, Richard T.  
; APPLICANT: Lea, Isabel  
; TITLE OF INVENTION: Sperm Antigen Corresponding to a Sperm  
; TITLE OF INVENTION: Zona Binding Protein Autoantigenic Epitope  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenneth D. Sibley  
; STREET: P.O. Box 34009  
; CITY: Charlotte  
; STATE: No. 5820861th Carolina  
; COUNTRY: USA  
; ZIP: 28234  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/436,883B  
; FILING DATE: 08-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sibley, Kenneth D.  
; REGISTRATION NUMBER: 31,665  
; REFERENCE/DOCKET NUMBER: 5470-73C  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-420-2200  
; TELEFAX: 919-881-3175  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-436-883B-57

Alignment Scores:  
Pred. No.: 5 31e-18 Length: 28  
Score: 28.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 9.12% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x US-08-436-883B-57 (1-28)

QY 298 TTTGATCCAGCAATGGGAGTAAGTGAAGACGGCTTCTATACCAATCATGCATTC 357  
|||||  
Db 1 PheAspProAlaGluTrpGlySerLysValGluAspArgPheTyraAsnHisAlaPhe 20  
QY 358 GAGGAGCAACACCACTTGAAA 381  
|||||  
Db 21 GluGluGluProGluLys 28

## RESULT 13

US-08-436-772-52  
; Sequence 52, Application US/08436772  
; Patent No. 5814456  
; GENERAL INFORMATION:  
; APPLICANT: O'Rand, Michael G.  
; APPLICANT: Widgren, Esther E.  
; APPLICANT: Richardson, Richard T.  
; APPLICANT: Lea, Isabel  
; TITLE OF INVENTION: Sperm Antigen Corresponding to a Sperm  
; TITLE OF INVENTION: Zona Binding Protein Autoantigenic Epitope

; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenneth D. Sibley  
; STREET: P.O. Box 34009  
; CITY: Charlotte  
; STATE: No. 5814456th Carolina  
; COUNTRY: USA  
; ZIP: 28234  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/436,772  
; FILING DATE: 08-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sibley, Kenneth D.  
; REGISTRATION NUMBER: 31,665  
; REFERENCE/DOCKET NUMBER: 5470-73B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-881-3140  
; TELEFAX: 919-881-3175  
; INFORMATION FOR SEQ ID NO: 52:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 25 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-436-772-52

Alignment Scores:  
Pred. No.: 3 91e-15 Length: 25  
Score: 25.00 Matches: 25  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 8.14% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x US-08-436-772-52 (1-25)

QY 145 CCATTCTCCACACCCACTACCGAATTCACAAAGATTGGGAATCTTCTGAAGGCTG 204  
|||||  
Db 1 ProPheSerAsnThrHisTyraGileProGlnGlyPheGlyAsnLeuGluGlyLeu 20  
QY 205 ACACGGAGATTCTG 219  
|||||  
Db 21 ThrArgGluIleLeu 25

## RESULT 14

US-08-436-883B-52  
; Sequence 52, Application US/08436883B  
; Patent No. 5820861  
; GENERAL INFORMATION:  
; APPLICANT: O'Rand, Michael G.  
; APPLICANT: Widgren, Esther E.  
; APPLICANT: Richardson, Richard T.  
; APPLICANT: Lea, Isabel  
; TITLE OF INVENTION: Sperm Antigen Corresponding to a Sperm  
; TITLE OF INVENTION: Zona Binding Protein Autoantigenic Epitope  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenneth D. Sibley  
; STREET: P.O. Box 34009  
; CITY: Charlotte  
; STATE: No. 5820861th Carolina  
; COUNTRY: USA  
; ZIP: 28234  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/436,883B  
FILING DATE: 08-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sibley, Kenneth D.  
REGISTRATION NUMBER: 31,665  
REFERENCE/DOCKET NUMBER: 5470-73B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-420-2200  
TELEFAX: 919-881-3175  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-436-883B-52

Alignment Scores:  
Pred. No.: 3,91e-15 Length: 25  
Score: 25.00 Matches: 25  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 8.14% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x US-08-436-883B-52 (1-25)

QY 145 CCATCTCCACACCCACTACCAATCCACAGGATTTGGGATCTCTTGAGGGCTG 204  
Db 1 ProPheSerAsnThrHisTyrArgIleProGIndGlyPheGlyAsnLeuLeuGluGlyLeu 20  
QY 205 ACACGCGAGATTCGTG 219  
Db 21 ThrArgGluIleLeu 25

RESULT 15

US-08-436-772-60  
Sequence 60, Application US/08436772  
Patent No. 5814456  
GENERAL INFORMATION:  
APPLICANT: O'Rand, Michael G.  
APPLICANT: Widgren, Esther E.  
APPLICANT: Richardson, Richard T.  
APPLICANT: Lea, Isabel  
TITLE OF INVENTION: Sperm Antigen Corresponding to a Sperm  
TITLE OF INVENTION: Zona Binding Protein Autoantigenic Epitope  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenneth D. Sibley  
STREET: P.O. Box 34009  
CITY: Charlotte  
STATE: No. 5814456th Carolina  
COUNTRY: USA  
ZIP: 28234  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/436,772  
FILING DATE: 08-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sibley, Kenneth D.  
REGISTRATION NUMBER: 31,665  
REFERENCE/DOCKET NUMBER: 5470-73B  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-881-3140  
TELEFAX: 919-881-3175  
INFORMATION FOR SEQ ID NO: 60:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-436-772-60

Alignment Scores:  
Pred. No.: 2,61e-11 Length: 21  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 6.84% Indels: 0  
DB: 2 Gaps: 0

AF334735 (1-954) x US-08-436-772-60 (1-21)

QY 481 GTTGCTGCTGTCAAAATCCAGCTCCCTCCGGGACACATAGCCAGAGGAGGCAAG 540  
Db 1 ValAlaAlaValIleGlnAlaAlaPheArgGlyHisIleAlaArgGluGluAlaLys 20  
QY 541 AAA 543  
Db 21 Lys 21

Search completed: February 15, 2005, 02:31:47  
Job time : 39.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 15, 2005, 02:28:57 ; Search time 147 Seconds  
(without alignments)  
4241.070 Million cell updates/sec

Title: AF334735

Perfect score: 307

Sequence: 1 TCGCCCTTCGCGCGCG.....AAAAAAAAAAAAAAAAAAAA 954

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1376875 seqs, 326749119 residues

Word size: 1

Total number of hits satisfying chosen parameters: 2645598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-Q=/cgn2\_1/USFTO.spool\_p/AF334735/runat\_14022005.160241.19433/app.query.fasta.1.1095  
-DB=Published Applications AA -QFWT=Fastan -SUFFIX=Oligo.rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=quality -THR MIN=1  
-ALIGN=15 -MODE=LOCAL -OUTFWT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USRR=AF334735 @CNG 1.1 199 @runat 14022005.160241.19433  
-NCFU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pcp.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pcp.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pcp.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pcp.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pcp.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pcp.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pcp.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pcp.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pcp.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pcp.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pcp.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pcp.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pcp.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pcp.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pcp.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pcp.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pcp.\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pcp.\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	11	3.6	157	15	US-10-424-599-166824	Sequence 166824,
2	11	3.6	236	15	US-10-424-599-162808	Sequence 162808,
3	10	3.3	113	16	US-10-437-963-121776	Sequence 121776,
4	9	2.9	63	15	US-10-424-599-211183	Sequence 211183,
5	9	2.9	104	15	US-10-424-599-174377	Sequence 174377,
6	9	2.9	105	16	US-10-437-963-162142	Sequence 162142,
7	9	2.9	126	15	US-10-424-599-247560	Sequence 247560,
8	9	2.9	137	15	US-10-424-599-249014	Sequence 249014,
9	9	2.9	202	15	US-10-424-599-249011	Sequence 249011,
10	9	2.9	592	14	US-10-100-294A-36	Sequence 36, Appl
11	9	2.9	602	15	US-10-104-047-3092	Sequence 3092, Ap
12	9	2.9	860	15	US-10-341-434-47	Sequence 47, Appl
13	9	2.9	860	15	US-10-416-592-2	Sequence 2, Appl
14	9	2.9	6620	15	US-10-080-334-290	Sequence 290, App
15	9	2.9	6620	16	US-10-408-765A-2291	Sequence 2291, Ap
16	9	2.9	7968	13	US-10-077-130-5	Sequence 5, Appl
17	9	2.6	23	10	US-09-986-480-311	Sequence 311, App
18	8	2.6	26	9	US-09-764-847-702	Sequence 702, App
19	8	2.6	26	14	US-10-092-154-702	Sequence 702, App
20	8	2.6	28	15	US-10-424-599-184595	Sequence 184595,
21	8	2.6	34	15	US-10-424-599-277778	Sequence 277778,
22	8	2.6	34	16	US-10-437-963-146931	Sequence 146931,
23	8	2.6	37	9	US-09-764-869-777	Sequence 777, App
24	8	2.6	37	14	US-10-091-504-777	Sequence 777, App
25	8	2.6	37	15	US-10-227-577-777	Sequence 777, App
26	8	2.6	38	16	US-10-437-963-105367	Sequence 105367,
27	8	2.6	39	15	US-10-424-599-208544	Sequence 208544,
28	8	2.6	39	15	US-10-424-599-247879	Sequence 247879,
29	8	2.6	40	16	US-10-437-963-109774	Sequence 109774,
30	8	2.6	42	16	US-10-437-963-186679	Sequence 186679,
31	8	2.6	43	16	US-10-437-963-156489	Sequence 156489,
32	8	2.6	44	15	US-10-424-599-195083	Sequence 195083,
33	8	2.6	44	16	US-10-291-226-91	Sequence 91, Appl
34	8	2.6	44	16	US-10-437-963-131469	Sequence 131469,
35	8	2.6	45	16	US-10-437-963-201534	Sequence 201534,
36	8	2.6	46	15	US-10-424-599-248344	Sequence 248344,
37	8	2.6	46	16	US-10-291-226-96	Sequence 96, Appl
38	8	2.6	46	16	US-10-437-963-154807	Sequence 154807,
39	8	2.6	47	9	US-09-814-122-48	Sequence 48, Appl
40	8	2.6	47	15	US-10-424-599-177226	Sequence 177226,
41	8	2.6	47	15	US-10-649-857-48	Sequence 48, Appl
42	8	2.6	48	15	US-10-424-599-170716	Sequence 170716,
43	8	2.6	48	15	US-10-424-599-254354	Sequence 254354,
44	8	2.6	48	15	US-10-424-599-256282	Sequence 256282,
45	8	2.6	49	15	US-10-424-599-254498	Sequence 254498,

ALIGNMENTS

RESULT 1  
US-10-424-599-166824  
; Sequence 166824, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 166824  
; LENGTH: 157  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_121657C.1.pcp  
US-10-424-599-166824



Alignment Scores: 157  
Pred. No.: 0.277 Length: 11  
Score: 11.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.58% Indels: 0  
DB: 15 Gaps: 0

AF334735 (1-954) x US-10-424-599-166824 (1-157)

QY 484 GCTGCTGTCAAATCCAAGCTGCTTCGGGGA 516  
Db 14 AlaAlaVallyslleGlnAlaAlaPheArgGly 24

## RESULT 2

US-10-424-599-162808  
; Sequence 162808, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 162808  
; LENGTH: 236  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_118033C.1.pbp  
US-10-424-599-162808

Alignment Scores: 236  
Pred. No.: 0.259 Length: 11  
Score: 11.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.58% Indels: 0  
DB: 15 Gaps: 0

AF334735 (1-954) x US-10-424-599-162808 (1-236)

QY 484 GCTGCTGTCAAATCCAAGCTGCTTCGGGGA 516  
Db 64 AlaAlaVallyslleGlnAlaAlaPheArgGly 74

## RESULT 3

US-10-437-963-121776  
; Sequence 121776, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazur, Brad  
; APPLICANT: Li, Ping  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 121776  
; LENGTH: 113

; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_24768C.1.pbp  
US-10-437-963-121776

Alignment Scores: 113  
Pred. No.: 2.67 Length: 10  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.26% Indels: 0  
DB: 16 Gaps: 0

AF334735 (1-954) x US-10-437-963-121776 (1-113)

QY 24 TTTTITTTTTTTTAAAGAAAAACGGTTA 53  
Db 23 PhePhePhePheLeuLysLysLysArgLeu 32

## RESULT 4

US-10-424-599-211183  
; Sequence 211183, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 211183  
; LENGTH: 63  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_32724C.1.pbp  
US-10-424-599-211183

Alignment Scores: 63  
Pred. No.: 26.9 Length: 9  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.93% Indels: 0  
DB: 15 Gaps: 0

AF334735 (1-954) x US-10-424-599-211183 (1-63)

QY 24 TTTTITTTTTTTTAAAGAAAAACGG 50  
Db 20 PhePhePhePheLeuLysLysLysArg 28

## RESULT 5

US-10-424-599-174377  
; Sequence 174377, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 174377

FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 247560  
LENGTH: 126  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_65576C.1.pap  
US-10-424-599-174377

Alignment Scores:  
Pred. No.: 24.7 Length: 104  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.93% Indels: 0  
DB: 15 Gaps: 0

AF334735 (1-954) x US-10-424-599-174377 (1-104)

QY 27 TTTTITTTTAAAGAAAAACGGTAA 53  
DB 10 PhePhePheLeuLysLysLysArgLeu 18

## RESULT 6

US-10-437-963-162142

Sequence 162142, Application US/10437963  
Publication No. US20040123343A1

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

APPLICANT: Wu, Wei

APPLICANT: Boukharov, Andrey A.

APPLICANT: Barbazuk, Brad

APPLICANT: Li, Ping

TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 38-21(53221)B

CURRENT APPLICATION NUMBER: US/10/437,963

CURRENT FILING DATE: 2003-05-14

NUMBER OF SEQ ID NOS: 204966

SEQ ID NO 162142

LENGTH: 105

TYPE: PRT

ORGANISM: Oryza sativa

FEATURE:

OTHER INFORMATION: Clone ID: PAT\_MRT4530\_61262C.1.pap  
US-10-437-963-162142

Alignment Scores:  
Pred. No.: 24.7 Length: 105  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.93% Indels: 0  
DB: 16 Gaps: 0

AF334735 (1-954) x US-10-437-963-162142 (1-105)

QY 928 ATCCCCCAAAAAAAAAAAAAA 954  
DB 87 IleProGlnLysLysLysLysLys 95

## RESULT 7

US-10-424-599-247560

Sequence 247560, Application US/10424599  
Publication No. US20040031072A1

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 247560  
LENGTH: 126  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_65576C.1.pap  
US-10-424-599-247560

Alignment Scores:  
Pred. No.: 23.9 Length: 126  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.92% Indels: 0  
DB: 15 Gaps: 0

AF334735 (1-954) x US-10-424-599-247560 (1-126)

QY 329 TCTACTTACTCCCCCATCTGCTGA 303

DB 16 SerThrLeuLeuProHisSerAlaGly 24

## RESULT 8

US-10-424-599-249014

Sequence 249014, Application US/10424599  
Publication No. US20040031072A1

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 38-21(53223)B

CURRENT APPLICATION NUMBER: US/10/424,599

CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ ID NOS: 285684

SEQ ID NO 249014

LENGTH: 137

TYPE: PRT

ORGANISM: Glycine max

FEATURE:

NAME/KEY: unsure

LOCATION: (1)..(137)

OTHER INFORMATION: unsure at all Xaa locations

FEATURE:

OTHER INFORMATION: Clone ID: PAT\_MRT3847\_66890C.1.pap  
US-10-424-599-249014

Alignment Scores:  
Pred. No.: 23.6 Length: 137  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.93% Indels: 0  
DB: 15 Gaps: 0

AF334735 (1-954) x US-10-424-599-249014 (1-137)

QY 487 GCTGTCAAAATCCAAAGCTTCGGG 513

DB 9 AlaValLysIleGlnAlaAlaPheArg 17

## RESULT 9

US-10-424-599-249011

Sequence 249011, Application US/10424599  
Publication No. US20040031072A1

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 249011  
; LENGTH: 202  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_66888C.1.pap  
US-10-424-599-249011

Alignment Scores:  
Pred. No.: 22.1 Length: 202  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.93% Indels: 0  
DB: 15 Gaps: 0

AF334735 (1-954) x US-10-424-599-249011 (1-202)

QY 487 GCTGTCAAAATCCAGCTGCTTCGG 513  
|||||  
DB 77 AlavallyaileGlnAlaAlaPheArg 85

## RESULT 10

US-10-100-294A-36  
; Sequence 36, Application US/10100294A  
; Publication No. US20030135877A1  
; GENERAL INFORMATION:  
; APPLICANT: HAERTEL, HEIKO A.  
; APPLICANT: MITTENDORF, VOLKER  
; APPLICANT: CHEN, RUOYING  
; APPLICANT: SHANK, KARIN J.  
; APPLICANT: RESKI, RALF  
; TITLE OF INVENTION: SUGAR AND LIPID METABOLISM REGULATORS IN PLANTS  
; FILE REFERENCE: 16313-0095  
; CURRENT APPLICATION NUMBER: US/10/100,294A  
; CURRENT FILING DATE: 2002-10-15  
; PRIOR APPLICATION NUMBER: 60/276,993  
; PRIOR FILING DATE: 2001-03-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 36  
; LENGTH: 592  
; TYPE: PRT  
; ORGANISM: Physcomitrella patens  
US-10-100-294A-36

Alignment Scores:  
Pred. No.: 18.5 Length: 592  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.93% Indels: 0  
DB: 14 Gaps: 0

AF334735 (1-954) x US-10-100-294A-36 (1-592)

QY 926 ACATCCCCCAAAAAAAAAAAAAAAAAA 952  
|||||  
DB 498 ThrSerProLysLysLysLysLys 506

## RESULT 11

US-10-104-047-3092  
; Sequence 3092, Application US/10104047  
; Publication No. US20030236392A1

; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: NO. US20030236392A1e1 full length cDNA  
; FILE REFERENCE: HI-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 3092  
; LENGTH: 602  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-104-047-3092

Alignment Scores:  
Pred. No.: 18.4 Length: 602  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.93% Indels: 0  
DB: 15 Gaps: 0

AF334735 (1-954) x US-10-104-047-3092 (1-602)

QY 926 ACATCCCCCAAAAAAAAAAAAAAAAAA 952  
|||||  
DB 148 ThrSerProLysLysLysLysLys 156

## RESULT 12

US-10-341-434-47  
; Sequence 47, Application US/10341434  
; Publication No. US20030215835A1  
; GENERAL INFORMATION:  
; APPLICANT: Origene Technologies  
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes  
; FILE REFERENCE: 9U 204 205 R1  
; CURRENT APPLICATION NUMBER: US/10/341,434  
; CURRENT FILING DATE: 2003-07-18  
; PRIOR APPLICATION NUMBER: US 60/348,164  
; PRIOR FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: US 60/348,119  
; PRIOR FILING DATE: 2002-01-15  
; NUMBER OF SEQ ID NOS: 238  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 47  
; LENGTH: 860  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-341-434-47

Alignment Scores:  
Pred. No.: 17.4 Length: 860  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.93% Indels: 0  
DB: 15 Gaps: 0

AF334735 (1-954) x US-10-341-434-47 (1-860)

QY 926 ACATCCCCCAAAAAAAAAAAAAAAAAA 952  
|||||  
DB 148 ThrSerProLysLysLysLysLys 156

## RESULT 13

US-10-416-592-2  
; Sequence 2, Application US/10416592  
; Publication No. US20040053291A1  
; GENERAL INFORMATION:  
; APPLICANT: TANG, Y. Tom  
; APPLICANT: WARREN, Bridget A.

APPLICANT: HAPALIA, April J.A.  
APPLICANT: ARVIZU, Chandra  
APPLICANT: LU, Dyung Aina M.  
APPLICANT: ELLIOTT, Vicki S.  
APPLICANT: XU, Yuming  
APPLICANT: LAL, Preeti G.  
APPLICANT: YUE, Henry  
TITLE OF INVENTION: NUCLEIC ACID-ASSOCIATED PROTEINS  
FILE REFERENCE: PF-0847 USN  
CURRENT APPLICATION NUMBER: US/10/416,592  
CURRENT FILING DATE: 2003-05-12  
PRIOR APPLICATION NUMBER: PCT/US01/43606  
PRIOR FILING DATE: 2001-11-13  
PRIOR APPLICATION NUMBER: US 60/248,404  
PRIOR FILING DATE: 2000-11-13  
PRIOR APPLICATION NUMBER: US 60/249,825  
PRIOR FILING DATE: 2000-11-16  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PERL Program  
SEQ ID NO 2  
LENGTH: 860  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20040053291A1 1933163CD1  
US-10-416-592-2

Alignment Scores:  
Pred. No.: 17.4 Length: 860  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.93% Indels: 0  
DB: 15 Gaps: 0

AF334735 (1-954) x US-10-416-592-2 (1-860)  
QY 926 ACATCCCCCAAAAAAAAAAAAAAAAAAAAAA 952  
Db 148 ThrSerProLysLysLysLysLysLys 156

RESULT 14  
US-10-080-334-290  
Sequence 290, Application US/10080334  
Publication No. US2004002584A1  
GENERAL INFORMATION:  
APPLICANT: Pena, Carol E. A.  
APPLICANT: Shimkets, Richard A  
APPLICANT: Li, Li  
APPLICANT: Shenoy, Suresh G  
APPLICANT: Kekuda, Ramesh  
APPLICANT: Spytak, Kimberly A.  
APPLICANT: Vernet, Corine A. M.  
APPLICANT: Malyankar, Uriel M  
APPLICANT: Guo, Xiaojia  
APPLICANT: Gusev, Vladimir Y  
APPLICANT: Casman, Stacie J  
APPLICANT: Boldog, Ferenc L  
APPLICANT: Furtak, Katarzyna  
APPLICANT: Tchernev, Velizar T  
APPLICANT: Patturajan, Meera  
APPLICANT: Gangolli, Esha A  
APPLICANT: Padigaru, Muralidhara  
APPLICANT: Liu, Xisohong  
APPLICANT: Baumgartner, Jason C.  
APPLICANT: Gerlach, Valerie  
APPLICANT: Spaderna, Steven K  
APPLICANT: Zerhusen, Bryan D  
TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of  
FILE REFERENCE: 21402-275  
CURRENT APPLICATION NUMBER: US/10/080,334

CURRENT FILING DATE: 2002-02-21  
PRIOR APPLICATION NUMBER: 60/270,523  
PRIOR FILING DATE: 2001-02-21  
PRIOR APPLICATION NUMBER: 60/322,712  
PRIOR FILING DATE: 2001-09-17  
PRIOR APPLICATION NUMBER: 60/311,980  
PRIOR FILING DATE: 2001-08-13  
PRIOR APPLICATION NUMBER: 60/330,307  
PRIOR FILING DATE: 2001-10-18  
PRIOR APPLICATION NUMBER: 60/278,796  
PRIOR FILING DATE: 2001-03-26  
PRIOR APPLICATION NUMBER: 60/281,521  
PRIOR FILING DATE: 2001-04-04  
PRIOR APPLICATION NUMBER: 60/276,677  
PRIOR FILING DATE: 2001-03-16  
PRIOR APPLICATION NUMBER: 60/311,595  
PRIOR FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: 60/270,220  
PRIOR FILING DATE: 2001-02-21  
PRIOR APPLICATION NUMBER: 60/274,295  
PRIOR FILING DATE: 2001-03-08  
PRIOR APPLICATION NUMBER: 60/318,526  
PRIOR FILING DATE: 2001-09-10  
PRIOR APPLICATION NUMBER: 60/286,548  
PRIOR FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/291,765  
PRIOR FILING DATE: 2001-05-17  
PRIOR APPLICATION NUMBER: 60/270,797  
PRIOR FILING DATE: 2001-02-23  
PRIOR APPLICATION NUMBER: 60/276,400  
PRIOR FILING DATE: 2001-03-16  
PRIOR APPLICATION NUMBER: 60/270,810  
PRIOR FILING DATE: 2001-02-23  
NUMBER OF SEQ ID NOS: 388  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 290  
LENGTH: 6620  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-080-334-290

Alignment Scores:  
Pred. No.: 12.3 Length: 6620  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.93% Indels: 0  
DB: 15 Gaps: 0

AF334735 (1-954) x US-10-080-334-290 (1-6620)  
QY 484 GCTGCTGCAAAATCCAGCTGCCTTC 510  
Db 4875 AlaAlaVallybIleGlnAlaAlaPhe 4883

RESULT 15  
US-10-408-765A-2291  
Sequence 2291, Application US/10408765A  
Publication No. US20040101874A1  
GENERAL INFORMATION:  
APPLICANT: Ghosh, Soumitra S.  
APPLICANT: Fahy, Eoin D.  
APPLICANT: Zhang, Bing  
APPLICANT: Gibson, Bradford W.  
APPLICANT: Taylor, Steven W.  
APPLICANT: Glenn, Gary M.  
APPLICANT: Warnock, Dale E.  
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
FILE REFERENCE: 660088.465  
CURRENT APPLICATION NUMBER: US/10/408,765A  
CURRENT FILING DATE: 2003-04-04  
NUMBER OF SEQ ID NOS: 3077

```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2291
; LENGTH: 6620
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2291

Alignment Scores:
Pred. No.:      12.3      Length:      6620
Score:          9.00      Matches:      9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      2.93%      Indels: 0
DB:              16      Gaps: 0

AF334735 (1-954) x US-10-408-765A-2291 (1-6620)

QY      484 GCTGCTGTCAAATCCAAGTCCTTC 510
      |||||
Db      4875 AlaAlaValLysIleGlnAlaAlaPhe 4883

```

Search completed: February 15, 2005, 02:51:50  
Job time : 160 secs